



Zhen Guo • Li Tan

FUNDAMENTALS AND APPLICATIONS
OF NANOMATERIALS

Fundamentals and Applications of Nanomaterials

For a listing of recent titles in the *Artech House Nanoscale Science and Engineering Series*,
turn to the back of this book.

Fundamentals and Applications of Nanomaterials

Zhen Guo
Li Tan



**ARTECH
HOUSE**

BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN-13: 978-1-59693-262-3

Cover design by Igor Valdman

© 2009 ARTECH HOUSE

685 Canton Street

Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

10 9 8 7 6 5 4 3 2 1

*To my wife Anne, sons Jonathan and Theodore Guo, and my parents
for their support and love.*

—Zhen

*To my family and friends for their efforts making my life at Lincoln interesting
and enjoyable.*

—Li

Contents

	Foreword	<i>xiii</i>
	Preface	<i>xv</i>
Part I	Fundamentals of Nanomaterials Science	1
1	Quantum Mechanics and Atomic Structures	3
1.1	Brief History of Quantum Mechanics	3
1.2	Photoelectric Effect and Duality Nature of Light	5
1.2.1	Photoelectric Effect	6
1.2.2	Einstein's Explanation	7
1.2.3	Duality of Light	9
1.3	Duality of Electrons	10
1.3.1	De Broglie's Hypothesis and Electrons as Waves	11
1.3.2	Time Independent Schrödinger Equation	12
1.3.3	Free Electrons	16
1.4	Electrons in Potential Well	17
1.4.1	1D Infinite Potential Well	17
1.4.2	3D Infinite Potential Well	19
1.5	Atomic Structure and the Periodic Table	21
1.5.1	The Hydrogen Atom	21
1.5.2	The Helium Atom	22
1.5.3	The Periodic Table	23

2	Bonding and Band Structure	25
2.1	Classic Atomic Bonding	25
2.2	Atomic Bonding in Molecules: LCAO Theory	26
2.2.1	Two-Atom Molecule	29
2.2.2	Three-Atom Molecule	31
2.2.3	Four-Atom Molecule	31
2.2.4	Six-Atom Molecule (Benzene Ring)	32
2.2.5	Many-Atom Molecule	34
2.3	Atomic Bonding in Crystalline Solids: Band Theory	35
2.3.1	Energy Band in Solids	36
2.3.2	Partially Filled Energy Band for Metals	37
2.3.3	Energy Band for Insulators and Semiconductors	38
2.4	Bonding and Band Structures in Nanocrystal Materials	40
2.4.1	Top-Down Method for Quantum Wells and Dots	41
2.4.2	Bottom-Up Method for Carbon-Based Nanocrystals	44
	References	47
3	Surface Science for Nanomaterials	49
3.1	Crystal Structure and Crystallography	49
3.1.1	Crystal Structures	49
3.1.2	Crystallography	50
3.1.3	Close-Packed Directions, Planes, and Structures	51
3.2	Surface Crystallography	55
3.2.1	Surface Structure for Close-Packed Structures	55
3.2.2	Surface Structure for BCC Structures	56
3.2.3	Surface Symmetry	57
3.3	Surface Energy	57
3.3.1	Crystallographically Preferred Surface	59
3.3.2	Wulff Constructions and Equilibrium Shape for Nanoparticles	59

3.4	Surface Reconfigurations	61
3.4.1	Surface Relaxation and Reconstructions	62
3.4.2	Adsorption	67
3.5	Surface Area and Surface Thermodynamics	68
3.5.1	Surface Area in Nanomaterials	69
3.5.2	Nanoparticle Nucleation	70
3.5.3	Wetting	71
	References	73
4	Nanomaterials Characterization	75
4.1	X-Ray Diffraction for Nanomaterials Characterization	75
4.1.1	X-Ray Diffraction and the Laue Method	76
4.1.2	Bragg's Law	77
4.1.3	X-Ray Diffraction in Nanomaterials	78
4.2	Electron Microscopy for Nanomaterials Characterization	80
4.2.1	Interaction Between Electron Beams and Solids	81
4.2.2	Transmission Electron Microscope (TEM)	83
4.2.3	Scanning Electron Microscope (SEM)	86
4.2.4	Scanning Probe Microscope (SPM)	87
4.3	Surface Analysis Methods	89
4.3.1	Auger Electron Spectroscopy (AES)	90
4.3.2	X-Ray Photoelectron Spectroscopy (XPS)	90
4.3.3	Secondary Ion Mass Spectroscopy (SIMS)	91
	References	91
Part II	Nanomaterials Fabrication	93
5	Thin-Film Deposition: Top-Down Approach	95
5.1	Thin-Film Deposition Mechanism	95
5.1.1	Homogeneous Film Growth Mechanisms	95
5.1.2	Heterogeneous Film Growth Mechanisms	97
5.2	Thin-Film Deposition Methods	99
5.2.1	Physical Vapor Deposition (PVD)	100
5.2.2	Chemical Vapor Deposition (CVD)	102
	Reference	106

6	Nanolithography: Top-Down Approach	107
6.1	Introduction	107
6.1.1	Parallel Replication	108
6.1.2	Serial Writing	110
6.2	Nanoimprint Lithography (NIL)	110
6.2.1	NIL Process	110
6.2.2	3D Patterning via NIL	111
6.2.3	Air Cushion Press	111
6.2.4	Sequential Embossing/Imprinting Lithography (SEIL)	113
6.3	AFM Lithography	114
6.3.1	Scratching and Nanoindentation	115
6.3.2	Nanografting	117
6.4	Polymer Pen Lithography (PPL)	119
6.5	Templated Self-Assembly of Block Copolymers	122
	References	126
7	Synthesis of Nanoparticles and Their Self-Assembly: Bottom-Up Approach	129
7.1	Synthesis of Nanoparticles	129
7.1.1	Coprecipitation	130
7.1.2	Sol-Gel Process	132
7.1.3	Microemulsions	135
7.1.4	Hydrothermal/Solvothermal Methods	136
7.1.5	Templated Synthesis	138
7.1.6	NPs of Organic Semiconductors	138
7.2	Self-Assembly of Nanoparticles	140
7.2.1	Hydrogen Bonding-Based Assembly	140
7.2.2	Electrostatic Assembly	141
7.2.3	Shape-Selective Assembly	142
7.2.4	Hydrophobic Assembly	144
7.2.5	Template-Assisted Assembly	144
7.2.6	Collective Properties of Self-Assembled Nanoparticles	146
7.3	Conclusion	147
	References	147

Part III	Nanomaterials Properties and Applications	151
8	Nanoelectronic Materials	153
8.1	Single-Electron Transistors (SETs)	153
8.1.1	Single-Electron Capacitor	154
8.1.2	Operating Principles for SETs	158
8.1.3	Quantum Effect on SETs	162
8.1.4	Fabrication of SETs	165
8.2	Carbon Nanotube-Based Nanoelectronic Device	170
8.2.1	Introduction to CNTs	170
8.2.2	Fabrications of CNTs	173
8.2.3	CNT-Based Devices	175
8.3	Spintronics	178
8.3.1	Operating Principles of Spintronics	178
8.3.2	Spintronics Devices	180
	References	184
9	Nano Biomaterials	187
9.1	Introduction	187
9.2	Biomimetic Nanotechnology	188
9.2.1	DNA Nanotechnology	188
9.2.2	Structural Biomimicry	191
9.3	Nanostructures in Biodiagnostics	198
9.3.1	Nanoparticle-Based Detection Methods	200
9.3.2	Nanowire- and Nanotube-Based Detection Methods	201
9.4	Nanostructures in Cells Study	202
9.4.1	Microarray Platform as a Research Tool	202
9.5	Tissue Engineering and Drug Delivery	205
	References	205
10	Nanostructural Materials	209
10.1	Nanograin-Sized Structural Materials	209
10.1.1	Why Grain Refinement?	210
10.1.2	General Approaches for Grain Refinement	210

10.2	Nanoindentation	213
10.2.1	Principles of Nanoindentation	214
10.2.2	In Situ Nanoindentation	214
10.3	Mechanical Instability of Nanostructures	219
10.3.1	Wrinkling of Thin Films	219
10.3.2	Buckling of Spheroidal Core/Shell Structures	224
10.3.3	Buckling of Nanobeams	225
10.3.4	Collective Buckling Model for Periodic Array of Nanostructures	229
	References	235
	About the Authors	239
	Index	241

Foreword

As one who has had the duty and privilege of teaching the elements of materials science to a few thousand engineering undergraduates over the last forty years, I am pleased to see this monograph on nanomaterials science appear, and am happy to write a brief Foreword to it.

The book is timely since the subject of nanomaterials has begun to coalesce into a definable subdiscipline of materials science with its own principles and features. This development has been some time in coming. To those of us in the more traditional areas of materials science, the importance of nanomaterials is nothing new. Metallurgists and ceramists have used and manipulated nanoscale precipitates and crystallographic defects for more than a century in the engineering of materials to control microstructure and properties. Specialists in electronic material have tailored structures and exploited behavior at the nanoscale for over half a century to achieve the breakthroughs in property and performance with which we are all familiar. Specialists in polymeric materials have, from the beginning, made materials that are assemblages of nanoscale molecules. So the idea of nanomaterials and their engineering exploitation is not novel at all.

Nor is it surprising that the level of effort devoted to advanced nanomaterials has grown so rapidly in recent years, as new methods of computation, characterization, and synthesis have been developed, and new engineering needs have proliferated. A major fraction of all of the research done on materials is now focused on those that are very small in size.

What is new is the idea that diverse nanomaterials have fundamental attributes in common besides their small size. And these common features reflect principles that, if mastered, offer a useful insight into the comparative behavior of very different nanomaterials, and a valuable framework for translating

engineering advances in the properties and processing of one class of nanomaterials into another.

Dr. Zhen Guo and Dr. Li Tan have taken on the task of surveying the broad field of nanomaterials from a perspective that stresses these underlying principles. They construct an outline of the field that includes applications from electronic materials through biological to structural materials, and methods from characterization through synthesis to technological application. At the same time they provide a reasonably comprehensive description of the particular classes of nanomaterials that have become important in modern materials engineering. The result is, in my opinion, not just a catalogue of interesting problems and applications in small systems, but a viable description of a subfield of materials science that promises to increase in importance as time goes on.

This monograph is one of the first to have taken this disciplinary approach to the broad subject of nanomaterials. It is targeted toward upper level students, who will find it a very useful text and reference.

*Dr. J. W. Morris, Jr.
Professor of material science and metallurgy, N.A.E.,
University of California, Berkeley
Berkeley, California
June 2009*

Preface

It has been exactly half a century since Dr. Richard Feynman (Nobel Laureate, 1964, Physics) gave the famous speech “There’s Plenty of Room at the Bottom,” in which he predicted that it is possible to put the entire 24 volumes of the *Encyclopedia Britannica* on the head of a pin. What was deemed impossible then is now reality. Fifty years later, no one would question that the twenty-first century is indeed a “Nano Era.” Driven by not only scientific curiosity, but also real life applications, nanomaterial offers vast opportunities from micro-electronics to aerospace, and from pharmaceuticals to clean energy. On the other hand, this rapidly growing multidisciplinary field makes it difficult for people to keep up to date except those entrenched in this area. Therefore, the superficially magic, sometimes even seemingly mysterious nanomaterial science poses significant challenges to universities and industries. They must figure out how to motivate the next generation of scientists and engineers to be better prepared for this new era of nanomaterials.

In order to answer this question, I started my part-time lecturing at San Jose State University about four years ago with several undergraduate and graduate level courses, including initiating and setting up a new nanomaterials course. My belief is that teaching is for not only transferring knowledge, but also for encouraging talented students to learn and explore science. In the course materials, I intentionally added many examples in which fundamental physics principles helped me solve real world problems. I invented an imaginary chip Spartan (SJSU’s Mascot) codeveloped by SJSU and Intel in 2012, and designed many of my homework assignments and exams around it. Furthermore, I gave the students a tour of the Intel Museum. Everyone was amazed by not only the “miracles” of microelectronic technology that integrates billions of nanometer scale transistors into a single chip used in our daily life, but also the

broad spectrum of scientific principles that make this a reality. In addition, I also invited many industry and academic experts to give speeches on different nanomaterials topics.

I observed, firsthand, that the interests and curiosities led to motivations and ultimately a deep understanding of principles and a high level of academic achievement. In the meantime, I also found myself inspired by those young brilliant minds and their determination and desire to learn new things. This was the primary motivation for me to write this textbook when Artech House approached me. My friend, Professor Li Tan at the University of Nebraska–Lincoln was also excited to hear about this opportunity and decided to delve in. He pulled out a few unique examples from his research group and fitted them under the umbrella of nanomaterial science.

The intention of this book is to introduce readers to a broad view of nanomaterials, from fundamental principles to fabrication techniques, and from characterization to applications. It is primarily intended as a textbook for seniors and/or graduate students in materials science and other related engineering fields. It can also serve as a reference book for professionals that are interested in this field. This book has the following unique features compared to other books in this field:

- This book is comprehensive and self-explanatory. Unlike most books in the nanomaterials field that focus mainly on current developments, this textbook covers fundamental principles, nanofabrications as well as nanomaterial properties and applications. Each part can be read independently or combined to create the complete book. Additionally, only a college level understanding of math, physics, and chemistry is required. We try to cover the most fundamental principles in this book, making it very suitable for a college level textbook or a handy reference book for professionals.
- This book covers a wide range of nanomaterials, such as nanodevice fabrications, nanomaterials characterization, and nanoelectronic, bio, and structural materials. The information included in this textbook accurately reflects the current state of nanomaterials, partially due to the authors diverse experience and broad interests in nanomaterials science and engineering. More importantly, we connected these developments with the fundamentals so that the readers will not only appreciate the magical aspects, but also understand the underlying principles.
- Whenever possible, we try to explain complicated problems with examples and physical concepts instead of mathematic formulas. The purpose of this book is not just to introduce the knowledge, but also to stimulate the interests of our young minds to nanomaterials. It is not magic or

mysterious; it is rather scientific. The authors have also listed many reference books and articles for those who would like to explore topics further.

This book is divided into three parts. Part I covers fundamentals of nanomaterials science, including Quantum Mechanics and Atomic Structures (Chapter 1), Bonding and Band Structures for Nanomaterials (Chapter 2), Surface Science (Chapter 3), and Nanomaterials Characterization (Chapter 4). Part II introduces nanofabrications, which covers both the top-down approach (Chapter 5: Thin Film Deposition and Chapter 6: Nanolithography) and the bottom-up approach (Chapter 7: Synthesis of Nanoparticles and Their Self-Assembly). Part III discusses nanomaterial properties and applications including Nanoelectronic Materials (Chapter 8), Nano Biomaterials (Chapter 9) and Nanostructural Materials (Chapter 10). Dr. Zhen Guo wrote Chapters 1–4, 5, and 9, and Sections 10.1, and 10.2. Dr. Li Tan wrote Chapters 6, 7, and 9, and Section 10.3.

Despite our every effort to make it perfect, there will be mistakes. If you find anything that needs to be corrected or clarified, please feel free to E-mail us at zguzguo@gmail.com or ltan4@unl.edu.

We are thankful to many friends and colleagues who have provided us with valuable suggestions to the course as well as this book. Specially, Professor J. W. Morris, Professor Andy Minor from UC Berkeley, Professor Gang-yu Liu from UC Davis, Professor Dave Mitlin from the University of Alberta, Professor Jiangyu Li from the University of Washington, Professor Jian Luo from Clemson University, and Professor Emily Allen from San Jose State University. In addition, Steve R. Smith, Ziguang Chen, Haojing Lin, and Dr. Chichao Yu have invested valuable time.

We also want to express our great appreciation to our publisher, Artech House for their continuous trust and encouragement, despite slipping deadlines on our side. Last but certainly not the least, we are deeply indebted to our families. Without their support and trust, this book is impossible.

Part I

Fundamentals of Nanomaterials Science

In this part, we are going to introduce some basic fundamental concepts of nanomaterials science, including quantum mechanics, atomic bonding, band structures, and surface physics. Some of the materials may be overlapping with college-level textbooks on materials science, but we think it is still worthwhile to introduce them again since they are so closely related to nanomaterials science. Hopefully, revisiting these topics will refresh the reader's memory. We will also focus closely on the implications of those concepts to nanomaterials science.

1

Quantum Mechanics and Atomic Structures

We certainly cannot start our discussion on nanomaterials without introducing quantum mechanics. Considered the most significant achievement in the last century, quantum mechanics is not only the foundation for modern physics theory, but also the major driving force for the so-called third wave of the Industrial Revolution—information technology, which is impacting everyone’s daily life. Computers, iPods, digital cameras, cell phones, to name just a few—none of these consumer electronic products would even be possible without quantum mechanics. In this chapter, we briefly introduce some basic knowledge on quantum mechanics and its implication for nanomaterials science.

1.1 Brief History of Quantum Mechanics

The 19th century saw the achievements of classic physics (i.e., Newtonian mechanics, Maxwell’s theory of electromagnetic fields) reach their peak. Many scientists believed that the framework of physics had already been discovered and that there were very few unsolved problems left for future generations. One good example is the discovery of the planet Pluto—people were able to predict not only the existence of Pluto but also its orbit and location based purely on calculated results from classic mechanics before its actual discovery.

As a consequence, determinism was popularly accepted among scientists. Laplace was one of the strong believers, as he expressed in the “Introduction to the Essai”:

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

However, as Chinese philosopher Confucius said 2,000 years ago, “There is no end for knowledge.” The thought that the exploration of physics was over was quickly been proved not only too optimistic but completely wrong. At the beginning of the 20th century, a series of experiment, including ones on black-body radiation and the photoelectric effect, put classic physics into crisis. The new experiments required new concepts and theories that were completely different from those of classic physics, which was at a loss to explain these phenomena.

Most historians in physics now attribute the beginning of development of quantum mechanics to Max Planck. While trying to give a theoretical derivation to the formula of blackbody radiation that he had guessed previously purely based on mathematics, he found he had to assume that total energy was not a continuous spectrum but rather made up of discrete energy elements—quanta of energy—an unprecedented step for quantum mechanics. Although Planck’s formula fits experimental data very well over a wide range of frequencies, his assumption of the quantum nature of radiation was hardly accepted. As a matter of fact, even Planck himself was puzzled by this assumption and wrote “Experience will prove whether this hypothesis is realized in nature.”

Several years later, Albert Einstein extended this theory to explain the photoelectric effect since traditional electromagnetic theory had a difficult time explaining experimental results. Inspired by Planck, he proposed the quantum theory of light in order to solve this dilemma. (More details are given in the next section.)

Since the success by Einstein explaining the photoelectric effect, many scientists in physics society started to accept the concept of quantum nature and began to use it to develop and explore the new atomic world. Bohr first proposed a model for the atomic structure of hydrogen atoms based on a discrete electron orbit and quantum energy. De Broglie suggested that not only light has both wave and particle duality as Einstein discovered with the photoelectric effect, but that all matter, including electrons, may have wave/particle duality. He also introduced the famous de Broglie relations to correlate the wavelength with a particle’s momentum. Schrodinger gave the electron its wave function

a few years later and successfully solved the problem for hydrogen atoms using wave mechanics.

The centerpiece of quantum mechanics is Heisenberg's uncertainty principle and the statistical meaning of the wave function. The uncertainty principle stated that one cannot accurately measure position x and momentum p of a particle at the same time because the process of measuring the position disturbs the particle's momentum. The statistical meaning of wave function meant that "one does not have an answer to the question, what is the state of an electron? But only to the question, how probable is it that an electron is going to occupy this state" (Born). It was the measurement that forced electron's wave function to collapse to a specific state.

Based on the principles above, quantum mechanics completely overthrew determinism. Even if one knows all of the current positions and forces, or everything in quantum mechanics, one still cannot determine all of its features. The impact to the physics community was so enormous that even Einstein had his personal doubts. He tried to prove that the reason quantum mechanics could not determine all of its features was not because "God is gambling" (Einstein's words) but because it was incomplete (hidden variable theory). Today, however, most physicists believe the other way—Mother Nature does exhibit uncertainty and statistical meaning in the atomic world.

Aside from the philosophical impact, quantum mechanics has been the driving force for the technological revolution in the last centuries, ranging from lasers to atomic radiation, from semiconductors and transistors to nanomaterials, from computers to medical instruments.

The purpose of this chapter is not to make you become a quantum mechanics expert but rather to introduce you to the basic quantum mechanics concepts that are applied in nanomaterials science.

1.2 Photoelectric Effect and Duality Nature of Light

Light was studied long before quantum mechanics. There are always two contradictory theories about light—theory of particles or theory of wave. For example, Isaac Newton believed that light was made up of numerous small particles in order to explain the fact that light travels in a straight line and reflects on the surface. However, this theory cannot explain the refraction and diffraction of light very well.

In the 19th century, people understood that light may be just another form of electromagnetic field that followed Maxwell's field equations. The wave character of the light was clearly demonstrated in the reflection, refraction, and diffraction of light. On the other hand, in order to explain how light, as a form of wave, transmits and travels through space, scientists had to proposed an

imaginary media, “aether,” just like the medium for mechanical wave propagation. However, this imaginary media has to have many conflicting properties in order to completely explain the properties of light. For example, it has to be fluid-like so it can occupy all the spaces. But it also has to be massless, transparent, incompressible, nondispersive, and have no viscosity. It will also have to be 1,000 times more rigid than steel in order to support the high frequency of light. By the end of the 19th century, the aether theory became so ironically magic that some scientists started to look for alternatives. The discovery of the photoelectric effect was the last straw that broke the camel’s back.

1.2.1 Photoelectric Effect

The photoelectric effect was first observed in the 19th century. As shown in Figure 1.1, the system consisted of a glass vacuum tube with both a metal cathode and anode sealed inside. The cathode and anode were connected through an ammeter to measure current and a voltage supply. It has been discovered that when the cathode is illuminated by incident light, the ammeter will record a current I .

Photoelectric emission obeys the following roles:

1. There exists a minimum frequency of incident light below which no current was detected. This frequency, named *threshold frequency* (ν_0), is determined by the material properties of cathode metal.
2. Once the frequency of incident light is above the threshold frequency (ν_0), photoelectric emission occurs. The current registered through the ammeter is directly proportional to the intensity of incident light.

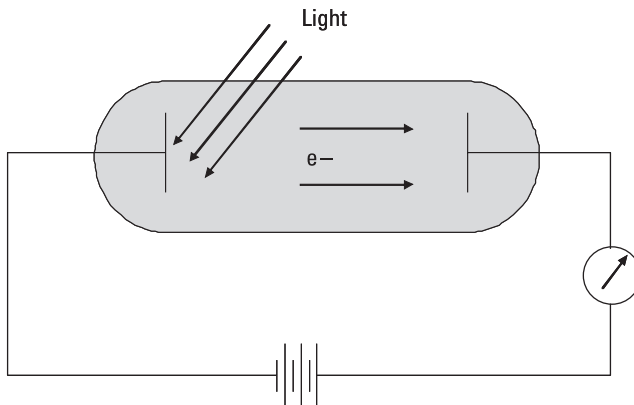


Figure 1.1 Illustration of the photoelectric effect.

3. The kinetic energy of emitted electrons from the metal cathode can be measured by applying reverse voltage until the current in the ammeter disappears. This reverse voltage is called *stopping potential*. It has been found that the stopping potential, and thus the kinetic energy of emitted electrons, is not related to light intensity but to the frequency of the incident light.
4. When increasing the forward bias, the current will increase. However, it will reach a saturation current, when all of the emitted electrons from the cathode are collected by the anode. Increasing the intensity of the incident light will increase the saturation current.

Because light has long been considered to be an electromagnetic wave, its energy should be proportional to the intensity as predicted by Maxwell theory in classic physics. However, in photoelectric experiments, the incident light energy and consequently the kinetic energy of emitted electrons seem to be proportional to the frequency, rather than the intensity of the incident light. For many years, scientists were puzzled by this dilemma until Einstein introduced the quantum nature of light in 1905.

1.2.2 Einstein's Explanation

Inspired by Max Planck's blackbody radiation theory, Einstein assumed that incident light consisted of many particles he called "light quanta" (later named *photons*). The energy of each individual photon can be expressed as follows:

$$E = h\nu \quad (1.1)$$

where h is Planck's constant and ν is the frequency of the light.

With this simple assumption, the photoelectric effect can be very easily interpreted as an interaction between incident photons and electrons inside metals. Each metal cathode has its own work function (ϕ_m), which is the energy required to emit a bonded electron from inside metal to vacuum. When the incident light shines on the metallic cathode, the photon is absorbed and its energy has been transferred to an electron. However, only when the incident photon energy is greater than the metal work function can the electron be emitted from metals. Therefore, there exists a threshold frequency for the incident light below which no electron will be emitted from the cathode and thus no current will be registered. This threshold frequency is characteristic of materials work function. The kinetic energy of emitted electrons is the difference between the energy carried by the incident light and the metal work function:

$$\text{KE} = h\nu - \phi_m = h\nu - h\nu_0 \quad (1.2)$$

where KE is kinetic energy of the emitted electron, h is Planck's constant, ν is the frequency of incident light, ν_0 is the threshold frequency, and ϕ_m is the materials work function.

When positive voltage is applied, more and more electrons are emitted from the cathode and collected at the anode. Hence, the current increases initially with an increase in positive voltage. However, once all electrons emitted from the cathode are collected by the anode, the current reaches a saturation level. This saturation current is a function of the intensity of the light. With higher intensity, the incident light consists of more photons and thus can emit more electrons, resulting in a higher saturation current.

When negative voltage is applied, the kinetic energy of emitted electrons will be overcome and eventually diminished when the negative voltage is large enough. This voltage value is called stopping voltage or V_0 as the current now stops flowing:

$$-eV_0 = \text{KE} = h\nu - h\nu_0 \quad (1.3)$$

Figure 1.2 illustrates the relationship between saturation current, incident light frequency/intensity, stopping potential, threshold frequency, and the cathode materials.

Exercise:

How do we determine experimentally through the photoelectric effect the threshold frequency ν_0 and thus work function ϕ_m of materials?

Using (1.3), we have $-eV_0 = \text{KE} = h\nu - h\nu_0$. For a given cathode material, we can measure different stopping potential V_0 under several incident light frequencies $[(V_0 \text{ versus } \nu \text{ as shown in Figure 1.2(b)}]$. This can be plotted in a

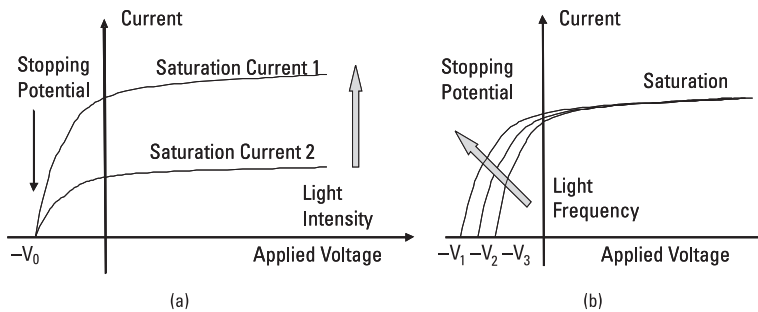


Figure 1.2 Illustration of current versus applied voltage for a given cathode material. (a) With no change of incident light frequency, the saturation current increases with an increase in incident light intensity. (b) Keeping the light intensity constant, the stopping potential increases with an increase in the incident light frequency.

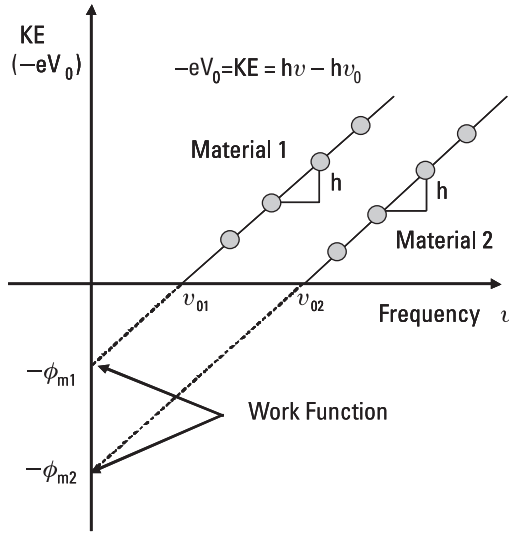


Figure 1.3 Obtaining the work function of cathode metals from the photoelectric effect.

linear relationship as shown in Figure 1.3. The slope of the line is h and the intercept on x axis is work function ϕ_m or $h\nu_0$.

1.2.3 Duality of Light

Although Einstein's theory explained the photoelectric effect very well, the idea of light quanta contradicted the long-held traditional belief that light is an electromagnetic wave following Maxwell's equation. So for a long time there was great resistance among physicists to this idea until Arthur Compton demonstrated the particle nature of light (called photons) through X-ray scattering (shown in Figure 1.4). He discovered that when an X-ray struck electrons, the frequency of the scattered X-ray (ν') is less than that of the incident X-ray (ν). Because we can also measure electron's kinetic energy (KE) and momentum after collision, it is straightforward to use the energy conservation law to find out whether the energy of an X-ray photon follows Einstein's photoelectric equation:

$$\text{KE} = h\nu - h\nu' \quad (1.4)$$

Now since the recoiling electron has momentum, we have to acknowledge the fact that the incident and the scattered X-ray photons also have momentum in order to follow the momentum conservation law. If we simplify this problem as a classical two-body elastic collision that follows conservation of both energy

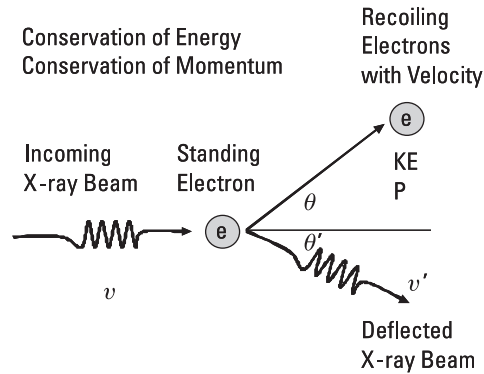


Figure 1.4 Compton's X-ray scattering experiment proved that an X-ray photon has energy that followed Einstein's photoelectric explanation and momentum that was related to wavelength.

and momentum, it can be shown that the momentum of photons is related to wavelength as well:

$$p = \frac{h}{\lambda} \quad (1.5)$$

From Compton's scattering experiment, we can see that light or photons, long considered as classic electromagnetic waves that obey Maxwell's theory, now not only have energy $h\nu$, but also a momentum p , both of which are typical characteristics for classic particles. This experiment showed Einstein's equation was accurate. Einstein's success with the photoelectric effect helped to prop up the emerging concept of duality. Light, traditionally considered to be an electromagnetic wave, exhibits the characteristics of waves and particles at different times. This duality nature was soon discovered in electrons, a typical particle, as we discuss in the next section.

1.3 Duality of Electrons

Unlike light, ever since electrons had been discovered in the 19th century, an electron had been considered to be a classical particle with its own position, speed, momentum, and energy that follows Newton's equation for classical mechanics. However, in the early 20th century, people started to realize the wave character of electrons, as discussed next.

1.3.1 De Broglie's Hypothesis and Electrons as Waves

Inspired by Einstein's equation and explanation of the photoelectric effect, French scientist Louis-Victor de Broglie proposed a hypothesis in 1924, claiming that all matter had wave-like nature (meaning wave/particle duality). He formulated a relationship between momentum, which represented particle characteristics, and wavelength, which represented its wave nature. He later completed his hypothesis in his Ph.D. thesis, which won him the Nobel Prize in 1929.

The de Broglie relation shows that the wavelength is inversely proportional to the momentum of a particle as follows:

$$\lambda = \frac{h}{p} \quad (1.6)$$

where λ is the particle's wavelength, h is Planck's constant, and p is the particle's momentum. This equation relates the wave-like and particle-like properties for any matter and thus serves as a bridge between the duality nature.

Although the hypothesis could be applied to all matters, it was first demonstrated in an experiment with electrons. As we know, interference and diffraction have long been recognized as distinguishing characteristics for a wave, such as light and X-ray. Not long after de Broglie formulated his hypothesis, scientists discovered that an electron beam could also show interference and diffraction with the right experimental setup as shown in Figure 1.5.

Figure 1.5(a) showed Young's double-slit experiment with electrons. When electrons are emitted from a filament and accelerated by forward bias passing a double slit, a pattern with alternating high- and low-intensity regions (as Young's fringes in the case of light) can be observed on the fluorescent screen at the back, indicating constructive and destructive interference. On the other

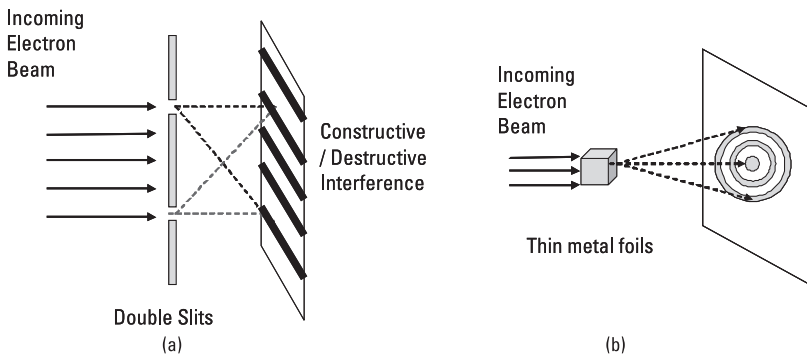


Figure 1.5 An electron exhibits the (a) double-slit interference and (b) thin metal foil diffraction that follows Bragg's relationship indicating the wave nature of electrons.

hand, if an electron beam is diffracted by a thin polycrystalline gold film, circular rings similar to X-ray diffraction predicted by Bragg's law are observed as shown in Figure 1.5(b).

The electron has long been considered to be a negatively charged particle that has its own energy, mass, momentum, charge, and speed, obeying traditional laws in classical mechanics, such as Newton's law, energy and momentum conservation, and the like. However, the fact that it showed interference and diffraction patterns demonstrated that it also had wave character and thus particle-wave duality.

1.3.2 Time Independent Schrödinger Equation

Since electron does exhibit wave nature, the next question will be whether it follows any wave equation just as traditional wave does. We shall discuss it in this section.

1.3.2.1 Traditional 1D Harmonic Wave Equation

By definition, a wave is a perturbation that propagates through space or a medium, transferring information, energy, and momentum. This propagation in physical space can be described mathematically using the wave equation. In the case of a 1D harmonic wave (as shown in Figure 1.6), the equation is

$$\Psi(x, t) = A \sin(kx - \omega t) = A \sin\left(\frac{2\pi}{\lambda}x - 2\pi f \times t\right) \quad (1.7)$$

where k is the wave number, ω is the angular velocity, λ is wavelength, and f is frequency. This wave equation can also be rewritten as

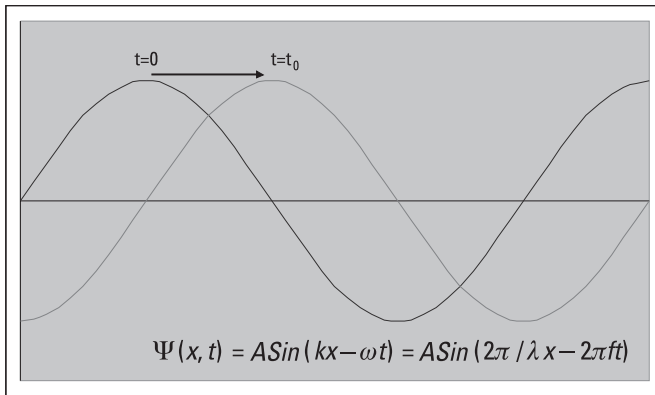


Figure 1.6 Illustration of standard 1D harmonic wave propagating along the x direction.

$$\Psi(x, t) = A \sin(kx - \omega t) = A \sin \left[k \left(x - \frac{\omega}{k} t \right) \right] = A \sin [k(x - V_p t)] \quad (1.8)$$

where V_p is phase velocity, which describes how fast the shape of the wave moves.

$$V_p = \frac{\omega}{k} = \lambda f \quad (1.9)$$

In the case of a harmonic sinusoidal curve, such as a 1D wave traveling through a rope with phase velocity V_p and amplitude Ψ , the wave equation is a differential equation

$$\frac{\partial^2 \Psi}{\partial t^2} = V_p^2 \frac{\partial^2 \Psi}{\partial x^2} \quad (1.10)$$

Another important velocity associated with wave is the group velocity V_g , which describes how fast the information and energy carried with the wave can propagate through the space:

$$V_g = \frac{\partial \omega}{\partial k} \quad (1.11)$$

The function of $\omega(k)$ is called the dispersion relationship. If ω is directly proportional to k , group velocity is the same as phase velocity.

We will use the equation introduced above to compare with wave function electrons in the following discussion of the Schrödinger equation.

1.3.2.2 1D Schrödinger Equations

As mentioned earlier, it has been discovered that electrons have particle-wave duality. As shown in Figure 1.5, electrons can exhibit interference and diffraction, just like light. In 1925, physicist Schrödinger analyzed what wave function electrons would follow if electrons indeed behave like a wave.

For a classical particle like an electron, the movement follows Newton's law. In the 1D case, the position of the electron at any given moment t can be described as $x(t)$. The velocity of that electron is $v = dx/dt$. If it is subjected to an external force, $F(x, t)$, it will follow Newton's second law $F = ma = m(dv/dt)$. On the other hand, in a conservative system the force can be described as derivative of potential energy and therefore:

$$F = -\frac{\partial V}{\partial x} = ma = m \left(\frac{dv}{dt} \right) = m \frac{d^2x}{dt^2} \Rightarrow m \frac{d^2x}{dt^2} = -\frac{\partial V}{\partial x} \quad (1.12)$$

This equation governs the movement of the electron. If we know the initial condition, that is, the electron's position at time zero $x(0)$ and potential energy $V(x)$, we should be able to determine the electron's position $x(t)$ at any given time.

However, things will be very different if we consider an electron to be a wave. As described in the previous section, a wave is usually described as a function of both space and time, $\Psi(x, t)$. Schrödinger proposed an equation that is analogous to the classic harmonic wave function as well as Newton's second law:

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + V(x, t) \Psi \quad (1.13)$$

This is called the 1D Schrödinger equation. Here \hbar is the Planck constant and equal to $h/2\pi$, m is electron mass, and V is the potential energy of the electron. The term $\Psi(x, t)$ is the wave format of an electron. Later on, Born proposed that the physical meaning of $|\Psi(x, t)|^2$ is the probability of finding an electron at location x and time t . Based on this assumption, $\Psi(x, t)$ has to obey the following rules:

1. $\Psi(x, t)$ has to be continuous and smooth in the space.
2. $|\Psi(x, t)|^2$ is the probability, so it has to be a real number. However, $\Psi(x, t)$ itself does not necessarily have to be real as $|\Psi(x, t)|^2 = \Psi^*(x, t) \Psi(x, t)$.

In the 3D case, the Schrödinger equation can be easily rewritten as

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V\Psi = -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} \right) + V\Psi \quad (1.14)$$

If the potential energy $V(x, t)$ is only a function of space, and does not change with time (in most cases, this assumption is valid), we can further simplify (1.13) to the 1D time-independent Schrödinger equation. Assuming

$$\Psi(x, t) = \psi(x) * f(t) \quad (1.15)$$

substituting this into (1.13), and rearranging the equation, we get:

$$i\hbar\psi\frac{\partial f}{\partial t} = -\frac{\hbar^2}{2m}f\frac{\partial^2\psi}{\partial x^2} + V(x)\psi f \Rightarrow i\hbar\frac{1}{f}\frac{\partial f}{\partial t} = -\frac{\hbar^2}{2m}\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} + V(x) \quad (1.16)$$

Because the left side of the equation is related to time and the right side is only related to space, the only possibility for this differential equation to work is if both sides equal a constant. Assume the constant is E . Then from the left side we have

$$i\hbar\frac{\partial f}{\partial t} = Ef \Rightarrow \frac{\partial f}{\partial t} = \frac{-iE}{\hbar}f \Rightarrow f(t) = \exp\left(\frac{-iE}{\hbar}t\right) = \exp(-i\omega t) \quad (1.17)$$

where $\omega = E/\hbar$. From the right side,

$$E = -\frac{\hbar^2}{2m}\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} + V(x) \quad (1.18)$$

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi = E\psi \Rightarrow \frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + (E - V(x))\psi = 0 \quad (1.19)$$

Equation (1.19) is called the 1D time-independent Schrödinger equation, which is widely used in quantum mechanics. On the right side of the equation we have a constant of E , which turns out to be the total energy of the particles. On the left side of the equation, we have a potential energy term $V(x)$ so the first term must be representing kinetic energy. If we assume:

$$\hat{P} = i\hbar\frac{\partial}{\partial x} \quad \text{and} \quad \hat{H} = \left(\frac{\hat{P}^2}{2m} + V\right) = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V\right) \quad (1.20)$$

then (1.19) can be rewritten as

$$\left(\frac{\hat{P}^2}{2m} + V\right)\psi = E\psi \Rightarrow \hat{H}\psi = E\psi \quad (1.21)$$

Here \hat{P} is the operator for momentum and \hat{H} is the Hamiltonian, which is the operator for energy. To obtain the energy from the Hamiltonian, we have to multiply the conjugate equation of wave function ψ^* and rewrite (1.21) to the following:

$$E = \frac{\int \psi \hat{H} \psi^* dV}{\int \psi \psi^* dV} \quad (1.22)$$

1.3.3 Free Electrons

Free electrons mean $V(x) = 0$ at any place. Hence, the Schrödinger equation can be written as

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{2mE}{\hbar^2} \psi = 0 \quad (1.23)$$

Because energy E is always positive, we can assume $k^2 = (2mE/\hbar^2)$ and rewrite the equation as

$$\frac{\partial^2 \psi}{\partial x^2} = -k^2 \psi \quad (1.24)$$

Solving this equation, we can obtain

$$\begin{cases} \psi_1(x) = A \exp(ikx) \\ \psi_2(x) = B \exp(-ikx) \end{cases} \quad (1.25)$$

The general solution is a linear combination of both solutions, expressed as

$$\psi(x) = A \exp(ikx) + B \exp(-ikx) \quad (1.26)$$

Considering the time term in (1.18), the overall wave function for free electrons is

$$\begin{cases} \Psi_1(x) = A \exp i(kx - \omega t) \\ \Psi_2(x) = B \exp i(-kx - \omega t) \end{cases} \quad (1.27)$$

The first solution is a traveling wave in the $+x$ direction, whereas the second is a traveling wave in the $-x$ direction.

For a free electron $V(x) = 0$, the kinetic energy is equal to total energy $\text{KE} = E = \hbar^2 k^2 / 2m$. Compared with classic mechanics where $\text{KE} = p^2 / 2m$, we can see the momentum is given as $p = \hbar k$. This is indeed a de Broglie relationship.

As we can see, the de Broglie relationship is a natural derivation from free electron behaviors based on the 1D time-independent Schrödinger equation. If we plot total energy E versus k , we will obtain a parabolic relationship as shown in Figure 1.7. This is called the dispersion relationship for free electrons.

1.4 Electrons in Potential Well

In this session, we will use Schrodinger equation discussed above to solve some simple quantum mechanics examples.

1.4.1 1D Infinite Potential Well

A 1D infinite potential well has zero potential inside the well and infinite potential outside [Figure 1.8(a)]. It can be written as

$$V(x) = \begin{cases} 0 & 0 \leq x \leq a \\ \infty & x > a, x < 0 \end{cases} \quad (1.28)$$

Therefore, outside the potential well, because $V(x)$ is infinitely high, the probability of finding an electron would be zero. So $\psi(x) = 0$ for $x > a$ or $x < 0$.

Inside the potential well, $V(x) = 0$. The electron is behaving like a free electron as we discussed in previous session. Therefore, the wave function is

$$\psi(x) = A \exp(ikx) + B \exp(-ikx) \quad \text{for } 0 \leq x \leq a \quad (1.29)$$

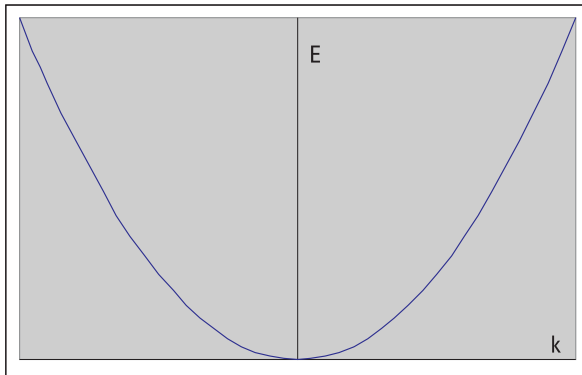


Figure 1.7 Dispersion relationship between energy E and wave number k for free electrons showing a parabolic curve.

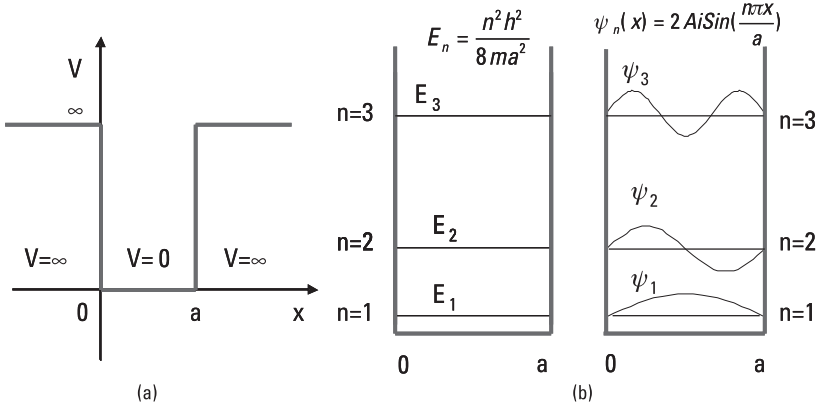


Figure 1.8 Schematic illustration of (a) a 1D infinite potential well and (b) its discrete energy level and wave function associated with it.

To fully determine the constants A and B , we need to determine the boundary conditions. Because $\psi^2(x)$ is described as the probability of finding an electron in a specific location, the wave function $\psi(x)$ will have to be smooth and continuous at all locations including the boundaries. Because outside the potential well $\psi(x) = 0$, it should also be zero at the boundaries. Therefore, we have $\psi(x) = 0$ at $x = 0$ and a . With these two boundary conditions, we can solve A and B .

When $x = 0$, $\psi(0) = A \exp(ik0) + B \exp(-ik0) = A + B = 0 \Rightarrow A = -B$. Therefore, the wave function $\psi(x) = A \exp(ikx) - A \exp(-ikx) = 2Ai \sin kx$. When $x = a$, $\psi(a) = 2Ai \sin ka = 0$. Note that A cannot be zero; otherwise, the wave function is zero everywhere, so we have $\sin ka = 0$, or $ka = n\pi$. This is equivalent to $k_n = n\pi/a$, $n = 1, 2, 3, \dots$, is an integer. Substituting it back into (1.29), we will get the following

$$\psi_n(x) = \begin{cases} 2Ai \sin\left(\frac{n\pi x}{a}\right) & 0 \leq x \leq a \\ 0 & x > a, x < 0 \end{cases} \quad (1.30)$$

We can also obtain the energy inside the 1D infinite well when $0 \leq x \leq a$,

$$E_n = -\frac{\hbar^2 k^2}{2m} = \frac{\hbar^2}{2m} \left(\frac{n\pi}{a}\right)^2 = \frac{n^2 \hbar^2 \pi^2}{2ma^2} = \frac{n^2 \hbar^2}{8ma^2} \quad (1.31)$$

where $n = 1, 2, 3, \dots$, is an integer, and we call it the quantum number. From the preceding equation, it is clear that the energy of an electron inside

an infinite potential well can only be at some finite and discrete levels instead of a continuous spectrum like a free electron. The *quantization* of wave function $\psi_n(x)$, wave number k_n , and energy E_n is a direct result of the Schrödinger equation and the fact that the wave function has to be continuous and smooth at the boundaries. Unlike classical mechanics, where a particle or object can take on any value of energy, in quantum mechanics, the energy of an electron can only have a quantized value. As a consequence, as shown in Figure 1.8, the wave function and wave number associated with energy level are also quantized.

On the other hand, the difference between two adjacent energy levels can be calculated as

$$\Delta E = E_{n+1} - E_n = \frac{(n+1)^2 \hbar^2}{8ma^2} - \frac{n^2 \hbar^2}{8ma^2} = \frac{(2n+1) \hbar^2}{8ma^2} \quad (1.32)$$

It is clear that when the size of potential well a increases, the difference between two adjacent energy levels is decreasing. When a is large enough, we can almost consider the energy level to no longer be discrete but continuous; at that time, quantum mechanics produces the same results as classical mechanics. Or in other words, when feature size is not at the atomic level but large enough, quantum mechanics predictions agree with classical results. This is called Bohr's correspondence principle.

Exercise:

Please calculate the ground state for an electron confined in a hydrogen atom and Harry Potter confined in the Chamber of Secrets. Compare the energy needed for them to jump from their ground state to the next energy level. Assume Harry Potter is 60 lb and the Secret Chamber is a 1D well at a size of 6 feet.

1.4.2 3D Infinite Potential Well

Like the 1D case, a 3D infinite potential well has zero potential inside the well and infinite potential outside:

$$V(x, y, z) = \begin{cases} 0 & 0 \leq x \leq a, 0 \leq y \leq b, 0 \leq z \leq c \\ \infty & \text{everywhere} \end{cases} \quad (1.33)$$

According to (1.14), the time-independent 3D Schrödinger equation can be written as

$$\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) + (E - V(x, y, z)) \psi = 0 \quad (1.34)$$

Similarly outside the potential well, because V is infinitely large, the probability of finding an electron is still zero. Inside the potential box, we can try to solve this equation using the separation of variables. We used this method to separate space and time in (1.15) to obtain the time-independent Schrödinger equation. Now we can use it again for the 3D infinite potential well case. Assume

$$\psi(x, y, z) = \psi_x(x) \times \psi_y(y) \times \psi_z(z) \quad (1.35)$$

Substituting this equation back to (1.34), we can obtain three separate differential equations, each one like in the case of the 1D infinite potential well. Since we have solved the 1D infinite well case, it is easy to see that the overall wave function is the product of three wave functions in the 1D case. Hence,

$$\psi(x, y, z) = A \sin\left(\frac{n_x \pi x}{a}\right) \sin\left(\frac{n_y \pi y}{b}\right) \sin\left(\frac{n_z \pi z}{c}\right) \quad (1.36)$$

$$\begin{aligned} E &= \frac{\hbar^2 k^2}{2m} = -\frac{\hbar^2 (k_x^2 + k_y^2 + k_z^2)}{2m} \\ &= \frac{h^2}{8m} \left[\left(\frac{n_x}{a}\right)^2 + \left(\frac{n_y}{b}\right)^2 + \left(\frac{n_z}{c}\right)^2 \right] \end{aligned} \quad (1.37)$$

Here n_x, n_y, n_z can be any integers except zero. The total energy of an electron confined in a 3D infinite potential well is determined by three quantum numbers— n_x, n_y, n_z associated with one wave function in each dimension. If the potential box is a cube, which means $a = b = c$, the total energy can be expressed as

$$E_{n_x, n_y, n_z} = \frac{h^2}{8ma^2} [n_x^2 + n_y^2 + n_z^2] \quad (1.38)$$

The ground state of energy for an electron in a 3D infinite potential box is

$$E_{111} = \frac{h^2}{8ma^2} [1 + 1 + 1] = \frac{3h^2}{8ma^2} \quad (1.39)$$

The second-level state would be

$$E_{211} = E_{121} = E_{112} = \frac{\hbar^2}{8ma^2} [1 + 1 + 4] = \frac{3\hbar^2}{4ma^2} \quad (1.40)$$

Because three different sets of quantum numbers or wave functions generate the same energy state, this is called degeneracy.

1.5 Atomic Structure and the Periodic Table

After the discussion in the previous section of the Schrödinger equation and electrons in potential wells, we are now ready to consider atomic structures. We begin with the hydrogen atom and then use it as a starting point for assembling the periodic table.

1.5.1 The Hydrogen Atom

Hydrogen is the lightest and also the simplest atom. It has one proton inside and one electron outside the nucleus. To the first order of approximation, we can consider both nucleus and electron as point charges. Therefore, the Coulombic potential energy can be described as

$$V(r) = \frac{-e^2}{4\pi\epsilon_0 r} \quad (1.41)$$

Substitute this into the 3D time-independent Schrödinger equation and we have

$$\nabla^2 \psi + \frac{2m_e}{\hbar^2} [E - V(r)] \psi = 0 \quad (1.42)$$

Notice that since the Coulombic potential is only a function of radius r , it may be appropriate to solve this 3D Schrödinger equation using (r, θ, ϕ) coordinates because the problem has a spherical symmetry. Similar to the way we separated variables (x, y, z) in Euclidian coordinates in the previous section, we can now express wave function $\psi(r, \theta, \phi)$ to be a product of a radius function $R(r)$ and a spherical function $Y(\theta, \phi)$. Without going through too many details, the wave function can be solved as follows:

$$\psi(r, \theta, \phi) = R(r) \times Y(\theta, \phi) = R_{n,l}(r) \times Y_{l,m}(\theta, \phi) \quad (1.43)$$

The wave function is determined by three quantum numbers—the principle quantum number n , the orbital angular momentum quantum number l ,

and the magnetic quantum number m_l . However, unlike the 3D potential well case in which three quantum numbers are independent of each other, in hydrogen atoms, it turns out that n , l , and m_l are related to each other and can only be assigned to certain values in order to achieve meaningful solutions for (1.43). Summarizing all of the allowable values for the three quantum numbers in hydrogen atoms, we get:

$$\begin{aligned} n &= 1, 2, 3, \dots \\ l &= 0, 1, 2, \dots, (n-1) \\ m_l &= -1, -(l-1), \dots, 0, \dots, (l-1), 1 \end{aligned} \quad (1.44)$$

Each possible wave function $\psi_{n,l,m_l}(r, \theta, \phi)$ with a combination of allowable n , l , m_l is a quantum state, sometimes referred as an *orbital*. The energy, on the other hand, is only a function of principle quantum numbers n for hydrogen atoms:

$$E_n = -\frac{me^4}{8\epsilon_0^2 h^2 n^2} \quad (1.45)$$

For example, for the ground state, $n = 1$, $l = 0$, and $m_l = 0$, the wave function is $\psi_{100}(r, \theta, \phi)$ and the energy at ground state is

$$E_{100} = E_1 = -\frac{me^4}{8\epsilon_0^2 h^2 l^2} = 2.18 \times 10^{-18} \text{ J} = 13.6 \text{ eV} \quad (1.46)$$

For the next energy level, there are several available states, $n = 2$, $l = 0$, $m_l = 0$; $n = 2$, $l = 1$, $m_l = -1$; $n = 2$, $l = 1$, $m_l = 0$; $n = 2$, $l = 1$, $m_l = 1$; or $\psi_{200}(r, \theta, \phi)$, $\psi_{21-1}(r, \theta, \phi)$, $\psi_{210}(r, \theta, \phi)$, $\psi_{211}(r, \theta, \phi)$. All four quantum states share the same energy (degenerate) since they have same principle quantum number n :

$$\begin{aligned} E_{200} = E_{21-1} = E_{210} = E_{211} = E_2 &= -\frac{me^4}{8\epsilon_0^2 h^2 2^2} \\ &= 5.45 \times 10^{-19} \text{ J} = 3.4 \text{ eV} \end{aligned} \quad (1.47)$$

1.5.2 The Helium Atom

Unlike a hydrogen atom, a helium atom has two electrons outside and two protons inside the nucleus. Thus, for any given electron, the potentials not only

depend on the Coulombic attraction between nucleus and electron as in hydrogen atoms, but also the repulsion from the other electron outside the nucleus. Now the nucleus has charge of $+2e$ and the potential energy of electron 1 can be described as a function of distance between itself and nucleus r_1 as well as the separation between two electrons r_{12} :

$$V_1(r_1, r_{12}) = \frac{-2e^2}{4\pi\epsilon_0 r_1} + \frac{e^2}{4\pi\epsilon_0 r_{12}} \quad (1.48)$$

Substituting this into the 3D time-independent Schrodinger equation, we can obtain the wave function and the energy associated with these quantum states just as we did for hydrogen atoms. Due to the presence of the interelectron potential term (r_{12}), the energy is now proportional to both n and l , instead of just n in the hydrogen atom case.

Traditionally, l values represent the spectroscopic notation. When $l = 0, 1, 2, 3$, we call them s (sharp), p (principle), d (diffuse), and f (fundamental). For example, for a hydrogen atom, the ground state is $n = 1, l = 0$, which is the $1s$ state. For $n = 2$, l can be either 0 or 1, so we have $2s$ and $2p$ states. However, for hydrogen atom, energy only depends on n so the $2s$ and $2p$ states are actually degenerate. For a He atom, since now energy depends on both n and l , the degeneracy between $2s$ and $2p$ is broken due to the interactions from the other electron outside the nucleus.

In general, the dependence of energy on n is higher than that on l . However, it may not be the case when l is large. For example, the energy for the $3d$ state ($n = 3, l = 2$) is actually higher than the $4s$ state ($n = 4, l = 0$), despite the fact that the $4s$ state has a high principle quantum number. So in terms of energy, we will have $1s < 2s < 2p < 3s < 3p < 4s < 3d < 4p < \dots$. This also holds true for heavier elements.

Figure 1.9 shows the relative energy level (not scaled) and its degenerated quantum states: $1s$ state: $n = 1, l = 0, m_l = 0$; $2s$ state: $n = 2, l = 0, m_l = 0$; and $2p$ state: $n = 2, l = 1, m_l = -1, 0, 1$ (three degenerated quantum states).

1.5.3 The Periodic Table

Now that we have energy levels for all of the quantum states, the question is how to arrange the electrons to form different elements accordingly. Besides the role of the lowest energy (electrons shall always occupy the quantum states of the lowest possible energy), there are a few more rules that nature follows when constructing the periodic table:

1. *Pauli exclusion principle*: Electrons cannot share the same exact quantum states. Therefore, any quantum states with a given combination of n ,

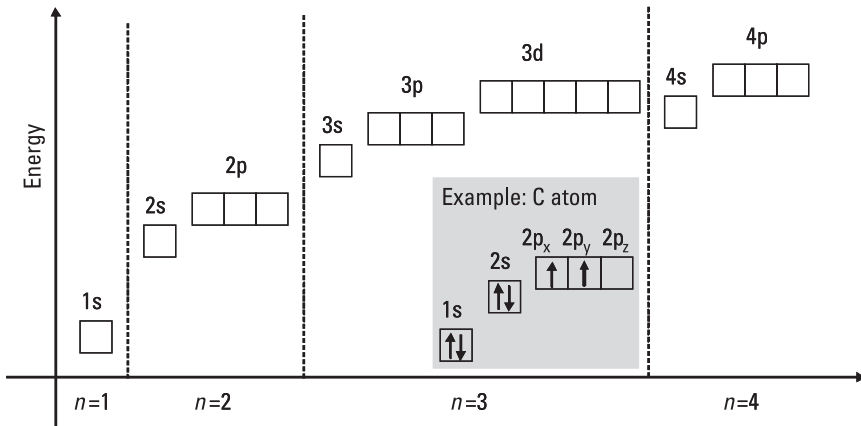


Figure 1.9 Energy level of atomic orbitals with different n , l , and m_l . The periodic table can be constructed based on the lowest energy level, the Pauli exclusion principle, and Hund's rule. An example is shown for the C atom.

l , m_l can only hold one pair of electrons, one spin up and one spin down. If we also describe spin as a quantum number using m_s and $m_s = 1/2, -1/2$ representing spin up and spin down, we then can state the Pauli exclusion principle simply as follows: For a given system, no two electrons shall have all four identical quantum numbers n , l , m_l , m_s , or no two electrons shall share the same quantum states.

2. *Hund's rule:* Just following the lowest energy and Pauli exclusion principle, we still cannot construct the entire periodic table. For example, when we try to fill $2p$ orbitals, which can contain a total of six electrons, we do not know which one to fill first. Deducted from spectroscopic studies, Hund's rule simply states that electrons in the same n and l orbitals would prefer their spins to be parallel (same m_s).

Now with those three rules, we can construct the periodic table. As shown in Figure 1.9, we use a blank box to represent an orbital with the same n , l , and m_l , and an arrow pointing up or down to represent electron with spin up or down. Using the lowest energy rule, we always fill up electrons from bottom (lowest energy) to top. Because electrons follow the Pauli exclusion principle, each box can only hold up to two electrons—one spin up and one spin down. When we start to fill $2p$, such as C atoms, the last electron will occupy $2p_y$ with spin parallel to the electron that was occupying $2p_x$. Therefore, entire the periodic table can be built based on these three principles.

2

Bonding and Band Structure

In the previous chapter, we discussed quantum mechanics and how atoms are formed. In this chapter, we discuss how atoms, as basic building blocks, form solids through atomic bonding and band structure.

2.1 Classic Atomic Bonding

When two atoms are brought close together, a bond will naturally develop between them. In general, the bonding seeks to balance the attractive Coulomb force that occurs between the nucleus in one atom and electrons in the other atom and the repulsive force that usually occurs between two nuclei or electrons of two atoms. The Coulomb force is the attraction between the positive and negative charge, and usually is proportional to $1/r$, where r is the distance between two atoms. On the other hand, the repulsive force is a short-range force, meaning that it increases very dramatically at short ranges, but diminishes very quickly as r increases.

Figure 2.1 shows both the attractive and repulsive force and their net force as a function of the interatomic distance r , assuming one atom at the origin. Attractive force is positive, and repulsive force is negative. When the interatomic distance is very large, two atoms are essentially separated with no bonding in between. So the force between two atoms is negligible. When we move the other atom closer to the origin, the attractive force starts to take effect first. As a result of this attractive force, the atoms want to move closer. As a consequence, the smaller the interatomic distance, the larger the attractive force and the lower the potential energy, indicating a stronger and more stable bond.

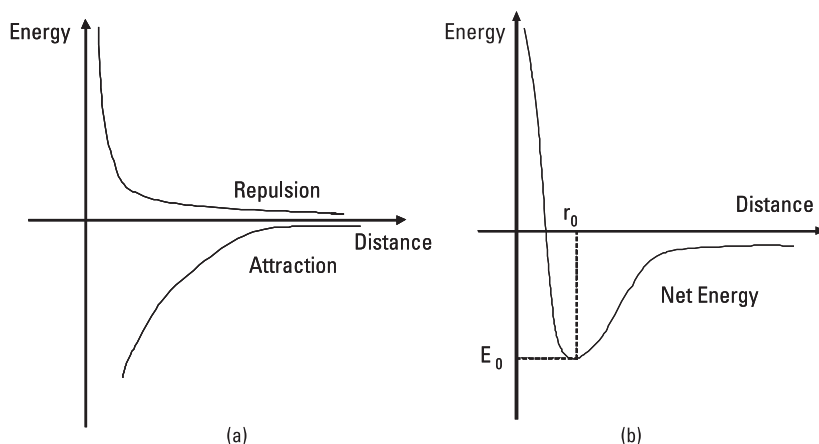


Figure 2.1 Classical atomic bonding model: (a) repulsion and attraction as a function of interatomic distance and (b) the overall energy of a molecule as a function of interatomic distance r showing a minimum at r_0 (bonding length) with energy of E_0 (bonding energy).

When atoms move too close, however, the electron clouds outside the nucleus start to overlap, leading to a repulsive force. Because the repulsive force is a strong short-range force, the net force decreases very fast with increasing interatomic spacing. At $r = r_0$, the net force becomes zero, and potential energy reaches its minimum point E_0 . So r_0 is the equilibrium interatomic spacing, called the *bonding length*, where bonding is the most stable. The potential energy E_0 is bonding strength or the energy required to break the bond and separate atoms. When atoms continue to move closer beyond r_0 , the repulsive force increases very dramatically, causing a great increase in potential energy.

This is the classical model for atomic bonding. It has successfully explained many phenomena such as thermal expansion and bulk elastic constant. However, this is only a qualitative model and it has limitations, especially in electronic structures of nanomaterials. In the next section, we introduce a quantitative approximation method called the *linear combination of atomic orbitals approximation* that is based on quantum mechanics and its application to nanomaterials. The next section may be a little bit math heavy, but readers can skip it and still reach the same conclusion.

2.2 Atomic Bonding in Molecules: LCAO Theory

The linear combination of atomic orbitals (LCAO) approximation is based on the assumption that when two or three or more atoms move closer to form a

molecule, the electrons from the individual atoms are now occupying molecular orbitals that can be expressed as a linear combination of atomic orbitals from each individual atom:

$$\Psi = \sum_{i=1}^n C_i \psi_i \quad (2.1)$$

where Ψ is the molecular orbital wave function, ψ_i is the atomic orbital wave function of the atom i in this n -atom molecule, and C_i is the contribution to molecular orbitals from this specific atomic orbitals.

Substituting (2.1) into the time-independent Schrödinger equation (1.22) and integrating it through volume dV , we have

$$\begin{aligned} E &= \frac{\int \Psi \hat{H} \Psi^* dV}{\int \Psi \Psi^* dV} = \frac{\int \sum_{i=1}^n C_i \psi_i \hat{H} \sum_{j=1}^n C_j \psi_j^* dV}{\int \sum_{i=1}^n C_i \psi_i \sum_{j=1}^n C_j \psi_j^* dV} \\ &= \frac{\sum_{i=1}^n C_i \sum_{j=1}^n C_j \int \psi_i \hat{H} \psi_j^* dV}{\sum_{i=1}^n C_i \sum_{j=1}^n C_j \int \psi_i \psi_j^* dV} \end{aligned} \quad (2.2)$$

Now we can define

$$H_{ij} = \int \psi_i \hat{H} \psi_j^* dV \quad \text{and} \quad S_{ij} = \int \psi_i \psi_j^* dV \quad (2.3)$$

Here S_{ij} is called the overlap integral between atom i and j in this n -atom molecule. The H_{ij} term represents the Coulomb integral when $i=j$ and resonance integral when $i \neq j$. We will discuss the detailed meaning of S_{ij} and H_{ij} very soon.

Then the equation above can be simplified as

$$E \sum_{i=1}^n \sum_{j=1}^n C_i C_j S_{ij} = \sum_{i=1}^n \sum_{j=1}^n C_i C_j H_{ij} \quad (2.4)$$

For any arbitrary molecular orbitals that consist of a linear combination of atomic orbitals, we will vary C_i to achieve the lowest energy for that specific

molecular orbital, which means $\partial E / \partial C_i = 0$. Taking the partial derivative of C_i from (2.4) above, we have

$$\begin{aligned} \frac{\partial E}{\partial C_i} \sum_{i=1}^n \sum_{j=1}^n C_i C_j S_{ij} + E \sum_{j=1}^n C_j S_{ij} &= \sum_{j=1}^n C_j H_{ij} \\ \Rightarrow \sum_{j=1}^n C_j (H_{ji} - ES_{ji}) &= 0 \end{aligned} \quad (2.5)$$

This set of equations can be used to solve C_j and can be rewritten in matrix form. In order for this set of equations to have real solution, we need the following necessary condition:

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \dots & H_{1n} - ES_{1n} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \dots & H_{2n} - ES_{2n} \\ \vdots & \vdots & & \vdots \\ H_{n1} - ES_{n1} & H_{n2} - ES_{n2} & \dots & H_{nn} - ES_{nn} \end{vmatrix} = 0 \quad (2.6)$$

This equation cannot be solved easily, so Hückel introduced a series of assumptions to simplify this equation:

Assumption 1:

All atoms in the molecule are identical. This means the molecule has a single element with n atoms. In this case, all Coulomb integrals are equal:

$$H_{ii} = \int \psi_i \hat{H} \psi_i^* dV = \alpha \quad (2.7)$$

Assumption 2:

The bonding only exists between neighboring atoms. This means the resonance integral is nonzero (assume it is β) between neighboring atoms and zero otherwise:

$$H_{ij} = \begin{cases} \int \psi_i \hat{H} \psi_j^* dV = \beta & \text{for } j = i \pm 1 \text{ (adjacent atoms)} \\ \int \psi_i \hat{H} \psi_j^* dV = 0 & \text{for other atoms} \end{cases} \quad (2.8a)$$

$$(2.8b)$$

Assumption 3:

The orbitals (wave functions) of each individual atom are orthogonal, meaning that overlap integrals between different atoms are zero or negligible:

$$\begin{aligned}
 S_{ii} &= \int \psi_i \psi_i^* dV = 1 \text{ for } i = j \text{ and} \\
 S_{ij} &= \int \psi_i \psi_j^* dV = 0 \text{ for } i \neq j
 \end{aligned}
 \tag{2.9}$$

Substitute (2.7), (2.8), and (2.9) into (2.6), we have

$$\begin{vmatrix}
 H_{11} - ES_{11} & H_{12} - ES_{12} & \dots & H_{1n} - ES_{1n} \\
 H_{21} - ES_{21} & H_{22} - ES_{22} & \dots & H_{2n} - ES_{2n} \\
 \vdots & \vdots & & \vdots \\
 H_{n1} - ES_{n1} & H_{n2} - ES_{n2} & \dots & H_{nn} - ES_{nn}
 \end{vmatrix} =$$

$$\begin{vmatrix}
 \alpha - E & \beta & 0 & \dots & 0 \\
 \beta & \alpha - E & \beta & \dots & 0 \\
 0 & \beta & \alpha - E & \dots & \vdots \\
 \vdots & \vdots & \vdots & \dots & \beta \\
 0 & 0 & \dots & \beta & \alpha - E
 \end{vmatrix} = 0
 \tag{2.10}$$

We will use this equation frequently throughout this section. The solution of this equation is called HMO (Hückel molecular orbitals).

2.2.1 Two-Atom Molecule

A two-atom molecule is the simplest form. Many gases have this form such as H_2 , N_2 , and O_2 . Because we studied the hydrogen atom extensively in Chapter 1, let's use H_2 as an example to study the two-atom molecule.

For a two-atom molecule, the molecular orbitals can be written as

$$\Psi = C_1 \psi_1 + C_2 \psi_2 \tag{2.11}$$

Substitute this into Schrödinger equation (2.5) and we have

$$\begin{cases}
 (\alpha - E) C_1 + \beta C_2 = 0 \\
 \beta C_1 + (\alpha - E) C_2 = 0
 \end{cases}
 \tag{2.12}$$

To solve the preceding equation, we will need

$$\begin{vmatrix}
 \alpha - E & \beta \\
 \beta & \alpha - E
 \end{vmatrix} = 0
 \tag{2.13}$$

The solutions are

$$\begin{cases} E = \alpha + \beta \Rightarrow C_1 = C_2 \Rightarrow \Psi = \psi_1 + \psi_2 \\ E = \alpha - \beta \Rightarrow C_1 = -C_2 \Rightarrow \Psi = \psi_1 - \psi_2 \end{cases} \quad (2.14)$$

Here α is the coulomb integral for each individual H atom, which is essentially the first ionization energy E_1 (-13.6 eV); β is the overlap integral between two adjacent H atoms so it represents bonding energy (~ -1 eV). The negative sign means that we will need to provide this amount of additional energy in order to either ionize the atom (α) or break the bond in the molecule (β). From (2.14), it is obvious that we have formed two molecular orbitals. One has energy $E = \alpha + \beta$ and wave function $\Psi = \psi_1 + \psi_2$. This is called a bonding molecular orbital, which has a lower energy than the other one, an antibonding molecular orbital that has $E = \alpha - \beta$ and $\Psi = \psi_1 - \psi_2$.

As shown in Figure 2.2, when two separated electrons, isolated at far left and far right, each with an energy level of α , come together to form an H_2 molecule, they will form both bonding molecular orbitals with energies of $(\alpha + \beta)$ and antibonding molecular orbitals with energies of $(\alpha - \beta)$. Because both α and β are negative, a bonding molecular orbital has the lower energy. Therefore, electrons from both hydrogen atoms occupy this state. After they form a molecule, the total energy is $2(\alpha + \beta)$ compared with the total energy of 2α for two isolated hydrogen atoms. The bonding energy is defined as the energy difference before and after the molecule is formed, in this case 2β . This is how H_2 molecules are formed.

In the next several sections, we first study three- and four-atom molecules using LCAO theory and then extend the discussion to a six or more atom system. We will not go through all of the mathematical details. Readers who are interested in this can solve these problems using linear algebra.

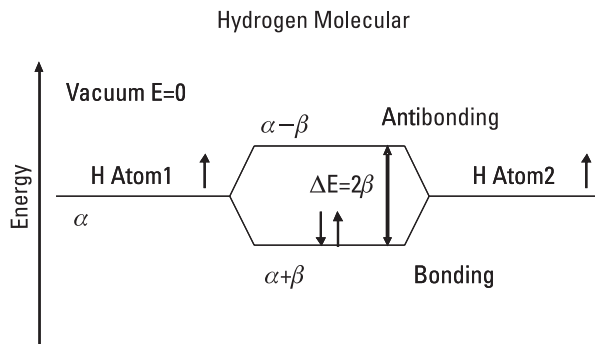


Figure 2.2 Schematic diagram showing the electron energy change of an H_2 molecule.

2.2.2 Three-Atom Molecule

Three atoms can form either a chain or a ring (triangle) molecule. For the chain structure, the bonding or electron overlap is formed between atoms 1–2 and 2–3, whereas for the ring structure, the overlap is formed not only as just mentioned but also between 3–1. The LCAO matrixes for these two types of molecules are described in (2.15a) for the chain and (2.15b) for the ring model:

$$\begin{vmatrix} \alpha - E & \beta & 0 \\ \beta & \alpha - E & \beta \\ 0 & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.15a)$$

$$\begin{vmatrix} \alpha - E & \beta & \beta \\ \beta & \alpha - E & \beta \\ \beta & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.15b)$$

The energy eigenvalues for these two equations can be solved through linear algebra. For the three-atom chain molecular model, $E_1 = \alpha + \sqrt{2}\beta$, $E_2 = \alpha$, $E_3 = \alpha - \sqrt{2}\beta$. For the three-atom ring molecular model, $E_1 = \alpha + 2\beta$, $E_2 = E_3 = \alpha - \beta$.

So if we still assume that each atom contributes one electron, when atoms are far away, the total energy of three electrons is 3α . When they move closer, atoms start to interact with each other to form molecules. Because both α and β are negative (see previous section for reason), E_1 is the lowest energy state, which can only contain one pair of electrons. The extra electron will have to go to an energy level of E_2 . For the three-atom chain molecular model, the total energy is $3\alpha + 2\sqrt{2}\beta$, whereas for the three-atom ring molecular model, the total energy is $3\alpha + 3\beta$. Because both α and β are negative numbers, the ring structure is energetically favorable for a three-atom molecule (shown in Figure 2.3).

2.2.3 Four-Atom Molecule

Just like the three-atom molecule, four atoms can also form either a chain or a ring (square) molecule. The LCAO matrix for these two types of molecular is described in (2.16a) for the chain model and (2.16b) for the ring model.

$$\begin{vmatrix} \alpha - E & \beta & 0 & 0 \\ \beta & \alpha - E & \beta & 0 \\ 0 & \beta & \alpha - E & \beta \\ 0 & 0 & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.16a)$$

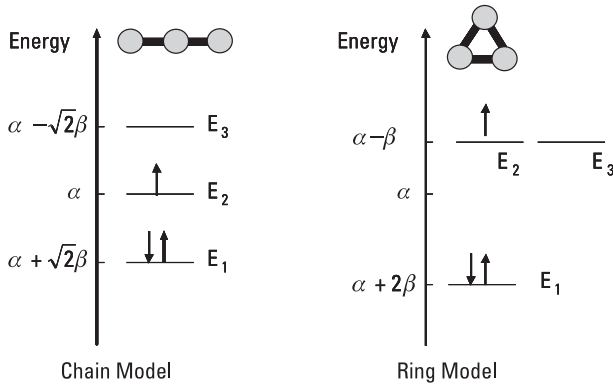


Figure 2.3 Energy level and molecular orbitals for three-atom molecule with chain and ring structures based on LCAO theory.

$$\begin{vmatrix} \alpha - E & \beta & 0 & \beta \\ \beta & \alpha - E & \beta & 0 \\ 0 & \beta & \alpha - E & \beta \\ \beta & 0 & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.16b)$$

The energy eigenvalues for these two equations can be solved through linear algebra. For chain molecules, $E_1 = \alpha + 1.618\beta$, $E_2 = \alpha + 0.618\beta$, $E_3 = \alpha - 0.618\beta$, $E_4 = \alpha - 1.618\beta$. For ring molecules, $E_1 = \alpha + 2\beta$, $E_2 = E_3 = \alpha$, $E_4 = \alpha - 2\beta$.

Similarly for a four-atom molecule, when atoms are far away, the total energy is 4α . When they move closer, electrons start to occupy the molecular orbit. In this case, four electrons will have to occupy both E_1 and E_2 . For the four-atom chain molecule model, the total energy is $4\alpha + 4.472\beta$. For the four-atom ring molecule model, the total energy is $4\alpha + 4\beta$. Because both α and β are negative, unlike the three-atom case, the four-atom molecule will form a chain structure rather than the rectangular ring.

As shown in Figure 2.4, in this case four electrons occupy the orbitals with energy levels of E_1 and E_2 , whereas E_3 and E_4 are completely empty. Therefore, we call orbitals with an energy level of E_2 the *highest occupied molecular orbitals* (HOMOs) and that of E_3 the *lowest unoccupied molecular orbitals* (LUMOs). Clearly, there is a gap between HOMO and LUMO (in this case, the gap is 1.236β). This gap is called the *HOMO-LUMO gap* (HLG).

2.2.4 Six-Atom Molecule (Benzene Ring)

Using the same methodology, we can also calculate six-atom molecules. In this case, we know the six-atom ring molecule is also called the benzene ring structure.

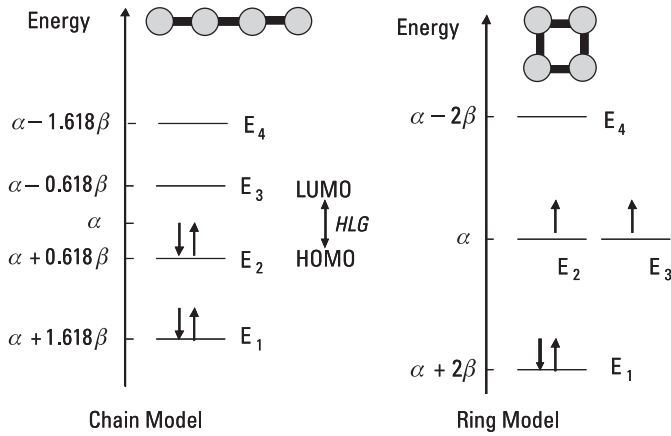


Figure 2.4 Energy level and molecular orbitals for four-atom molecule with chain and ring structures based on LCAO theory.

The chain structure is also commonly observed. The LCAO matrices for these two types of molecules are described in (2.17a) for the chain model and (2.17b) for the ring model.

$$\begin{vmatrix} \alpha - E & \beta & 0 & 0 & 0 & 0 \\ \beta & \alpha - E & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha - E & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha - E & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha - E & \beta \\ 0 & 0 & 0 & 0 & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.17a)$$

$$\begin{vmatrix} \alpha - E & \beta & 0 & 0 & 0 & \beta \\ \beta & \alpha - E & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha - E & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha - E & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha - E & \beta \\ \beta & 0 & 0 & 0 & \beta & \alpha - E \end{vmatrix} = 0 \quad (2.17b)$$

The energy eigenvalues for these two equations can be solved through linear algebra. For chain molecules, $E_{1,6} = \alpha \pm 1.802\beta$, $E_{2,5} = \alpha \pm 1.247\beta$, $E_{3,4} = \alpha \pm 0.445\beta$. For ring molecules, $E_1 = \alpha + 2\beta$, $E_2 = E_3 = \alpha + \beta$, $E_4 = E_5 = \alpha - \beta$, $E_6 = \alpha - 2\beta$.

Again for a six-atom molecule, when atoms are far away from each other, the total energy is 6α . When they move closer, electrons started to occupy molecular orbit. In this case, six electrons will have to occupy orbitals with energy levels of E_1 , E_2 , and E_3 . Therefore, for a six-atom chain molecular model, the total energy is $6\alpha + 7\beta$. For a six-atom ring molecular model, the total energy is $6\alpha + 8\beta$.

From calculations based on the LCAO theory discussed earlier, six atoms such as C atoms should always form a benzene ring structure. However, we know that LCAO theory is a simplified approximation that does not consider many other interactions, such as π bonding and interaction with other atoms such as C–H bonding. Hence in reality, both ring and chain structures are observed (Figure 2.5).

2.2.5 Many-Atom Molecule

We have calculated two-, three-, four-, and six-atom molecules. What we have observed is that the number of molecular orbital energy levels is equal to the number of atoms in the molecule. Also the energy levels are systematically distributed below and above α . On the other hand, the ground energy E_1 becomes lower and closer to $\alpha + 2\beta$ as the number of atoms in a many-atom chain molecule increases. Because there are enough energy states, all of the valence electrons occupy the energy states below α . We can keep adding the number of atoms and still get a similar matrix. Waser [1] calculated the conjugated hydrocarbon with unbranched chain structure of many atoms using

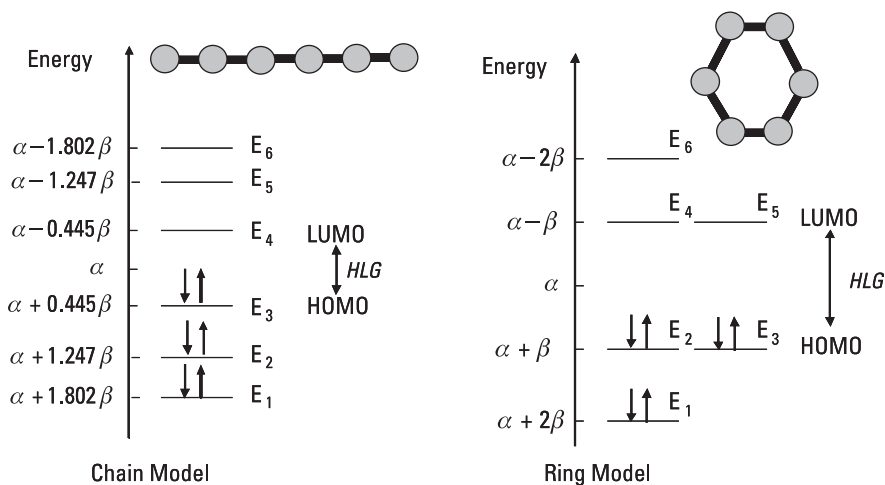


Figure 2.5 Energy level and molecular orbitals for six-atom molecule with chain and ring structures based on LCAO theory.

LCAO theory and HMO approximations. Figure 2.6 shows an example with 10 atoms. It is clear that all occupied orbitals have energy levels between $\alpha + 2\beta$ and α , whereas all empty states are between α and $\alpha - 2\beta$. Because all energy levels are confined to this range, when the total number of energy levels increases with the number of atoms in the molecule, the gap between adjacent energy levels becomes smaller and, eventually, when the number of atoms considered is large enough, we can consider it to be a continuous energy spectrum rather than a discrete energy level since the energy difference between the two energy levels is negligible. This is our introduction to conduction and valence bands, which we are going to discuss in detail in the next section.

2.3 Atomic Bonding in Crystalline Solids: Band Theory

In the previous section, we discussed atomic bonding in multiple-atom molecules. Using LCAO theory and linear algebra, we can obtain reasonably simple results for energy levels for those molecules. Although the model is oversimplified, the results, however, at least qualitatively match with experimental data.

On the other hand, in crystalline solids, there are $\sim 10^{23}$ atoms in a 3D lattice structure. LCAO theory based on linear algebra is no longer useful. Just

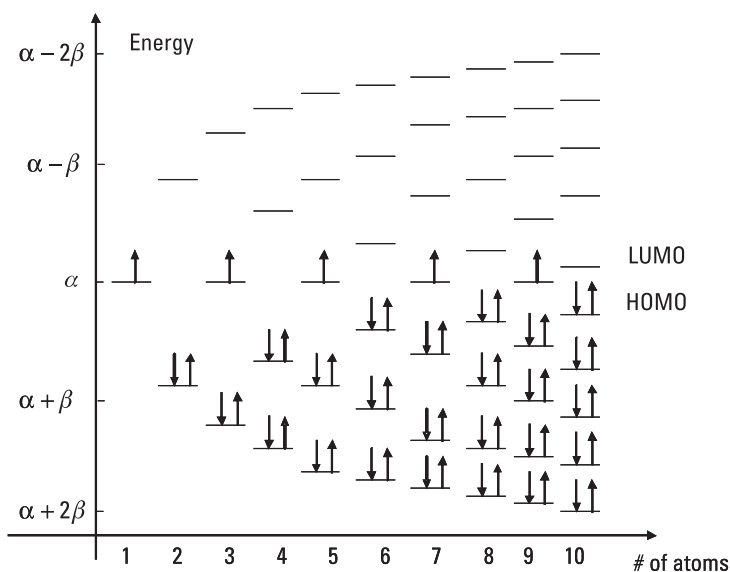


Figure 2.6 Schematic illustration of energy level in a multiatom molecule with chain structure, showing that electrons are mostly occupying the energy level below α and that a gap exists between HOMO and LUMO. (Adapted from [1].)

imagine trying to use it to solve such a big and sparse matrix, not to mention that the lattice is now 3D so we might need a tensor instead of a matrix. We have to use other methods to study the energy levels in crystalline solids. In this section, we introduce band theory and how to use it to study the electric properties of materials.

2.3.1 Energy Band in Solids

Using LCAO theory, we have calculated the energy levels of molecules with several atoms. The results show that each atom will contribute one molecular orbital, half of which is above and half below the energy level for isolated atoms α .

We can still use hydrogen atoms as an example. For two hydrogen atoms, as we discussed in the previous section, their $1s$ orbitals ($\alpha = -13.6$ eV) will split into two orbitals, one bonding orbital and one antibonding orbital. With one unpaired electron of each H atom, these two electrons will fill the bonding orbital, which has a lower energy, and leave the antibonding orbital empty. The net results are that the total energy of the H_2 molecule is lowered by 2β compared to that of two isolated H atoms.

If we have six atoms, it will have six orbitals in total, among which half are below and half are above α . If each atom contributed one unpaired electron to the molecule for bonding, the six unpaired electron will only occupy the three orbitals that have the lowest energy level. Therefore, all of the energy levels below α are filled, whereas those above are empty. Because an energy level is not continuous but discrete, we have thus defined HOMO, LUMO, and HLG as discussed in the previous session.

If we have N atoms in a solid where N is very large (for example, in metal Li, N is about 10^{23}), we can still use the results from the previous section for a qualitative discussion: N atoms will provide N orbitals for solids among which half are between $\alpha + 2\beta$ and α , the other half are between α and $\alpha - 2\beta$. Here, α and β are finite numbers that are independent of the number of atoms. However, now we have a large number of energy levels that need to be put into this finite range of energy. Because N is very large, essentially the energy levels are no longer discrete but continuous. This energy span containing continuous energy levels is called an *energy band*.

Each energy band in general corresponds to one energy level of an atomic orbital in an isolated atom. However, in some cases the energy band will overlap to form a continuous one or split into two separate ones. Between each band is the forbidden gap where no orbitals are allowed and thus no energy level exists. At absolute zero degree, electrons fill an energy band from the lowest energy level to the highest (HLG). Usually the highest energy band occupied by electrons is called the *valence band* (VB). The valence band can be either

fully occupied or partially filled. If the valence band is fully occupied, the next available energy band is called the *conduction band* (CB).

Next, we provide more details on energy bands using examples of metals, semiconductors, and insulators.

2.3.2 Partially Filled Energy Band for Metals

Let's start with the simplest metal Li. The Li atom has a fully occupied $1s$ orbital and a partially occupied $2s$ orbital. As shown in Figure 2.7, when Li atoms are isolated, the energy levels for the $1s$ and $2s$ orbitals are E_{1s} and E_{2s} , respectively. When we bring N atoms together to form a solid, the $2s$ level will split to N finite energy levels. From discussions based on LCAO theory in previous sections, all of those N energy levels now locate between $\alpha + 2\beta$ and $\alpha - 2\beta$. The maximum separation is $\sim 4\beta$ on the order of 5 to 10 eV. On the other hand, N is such a large number (on the order of 10^{23}) that the discrete energy levels in the range of $\alpha + 2\beta$ and $\alpha - 2\beta$ can be considered as a continuous *energy band*.

In this case, the $2s$ energy band consists of N orbitals, each of which can contain two spin-paired electrons. Thus, the $2s$ energy band can contain up to $2N$ electrons. However, in a Li solid, N atoms can contribute N $2s$ electrons. Those electrons can only occupy half of the N orbitals. As shown in Figure 2.7, the bottom half of the $2s$ energy band is filled completely, whereas the top half is empty. Therefore, a half-filled $2s$ energy band is formed in a Li solid from an isolated half-filled $2s$ atomic orbitals.

The electrical property of Li is directly determined by the electron configuration in this partially filled energy band. When a small electrical field V_0 is applied, the electron energy on the left side is raised a delta of eV_0 . Because the energy band is only partially filled, many empty orbitals with higher energy levels are available in the same band. Therefore, electrons on the left side have

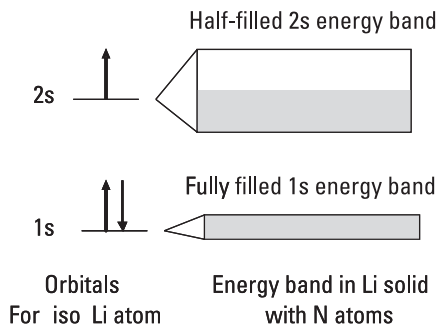


Figure 2.7 Schematic illustration of Li bandgap theory showing a partially filled $2s$ energy band, which leads to its conductive electric properties.

higher energy $E + eV_0$ compared to the right side and, thus, electrons will automatically move from the left side to the right to pass a current. This is why Li is conductive and can easily conduct current due to this partially filled energy band.

A similar discussion applies to Be. Beryllium has $2s^2$ orbitals. So when N Be atoms are brought together, the $2s$ orbitals now split into N orbitals that can contain $2N$ electrons, just as in the Li case. However, each Be has two electrons in $2s$ orbitals and thus N Be atoms will provide $2N$ electrons, just enough to fill the full $2s$ energy band. So at first glance, it seems that Be should have a fully occupied $2s$ energy band and a completely empty $2p$ band—but, then, what makes it conductive?

In reality, the empty $2p$ band, formed from $2p$ orbitals in isolated atoms, is now spreading so wide that it is actually overlapping with the $2s$ band. In other words, the β from $2s$ and $2p$ are large enough that it actually offsets the delta of α between $2s$ and $2p$ (as shown in Figure 2.8). Thus, those $2s$ and $2p$ bands are actually one continuous band, which provides empty energy levels for those $2N$ $2s$ orbitals. This partially filled energy band gives conducting electrons sufficient room to gain energy from the field and contribute to electric current. Solid Be, therefore, is a conductive metal.

2.3.3 Energy Band for Insulators and Semiconductors

For insulators and semiconductors, let's use diamond and silicon and carbon as examples. Both carbon and silicon belong to the IVA group, which has 4 valence electrons in its outer shell orbital. For a C atom, the electron configuration is $1s^2 2s^2 2p^2$. In theory, $1s$ and $2s$ orbitals are fully occupied and only $2p$

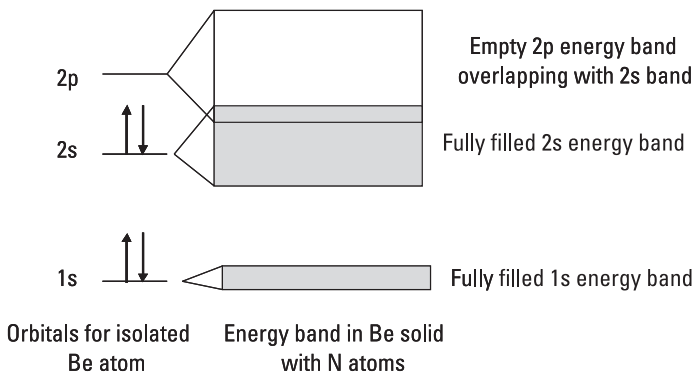


Figure 2.8 Schematic illustration of band structure of Be. Although it has a fully filled band from the $2s$ atomic orbital, it is still a conductive metal because its band from the $2p$ orbital is overlapping with the one from the $2s$, therefore creating a continuous and large band structure. As a consequence, Be is a conductive metal.

is the valance band. In reality, however, because the $2s$ and sp energy levels are so close, they are very likely mixing with each other and forming four hybrid sp^3 orbitals when many atoms are brought together to form a solid. These four hybrid sp^3 orbitals share the same energy level and have similar space configurations (Figure 2.9). Each orbital contains one electron so they are half filled and ready to form covalence bonds with their neighbors. Because all four orbitals are evenly distributed in the space, they have an angle of 108° with each other in tetrahedral directions. Therefore, when a C or Si atom bonds with its neighbors, it forms a tetrahedral structure with all four neighboring atoms at four corners sharing four covalence bonds with the C or Si atom in the center. This is called the diamond structure—both Si and diamond have this structure.

If we have a large number (N) of Si or C atoms, we will have a total of $4N$ orbitals and $4N$ valence electrons. At first glance, it seems like this should be a partially filled band structure, which will make materials conductive. But in reality we know that diamond is an insulator and Si is a semiconductor. What really happens is that when we put N atoms together and create bonds between adjacent atoms, bonding and antibonding orbitals will be formed, just like in the H_2 molecule case discussed in the previous section. The bonding orbitals have lower energy and the antibonding orbitals have energy higher than the isolated atom. Therefore, $4N$ orbitals will be split into two groups, half for bonding orbitals and half for antibonding orbitals, between which there is an energy gap where no states are allowed. As shown in Figure 2.9, the entire sp^3 in solid C or Si will split into two bands—one for bonding orbitals and one for antibonding orbitals, each containing $2N$ states. Because we have a total of N atoms and thus $4N$ valence electrons, those electrons will fully occupy the bonding orbitals while keep the antibonding orbitals completely empty. Here, the fully occupied energy band from bonding molecular orbitals is called the

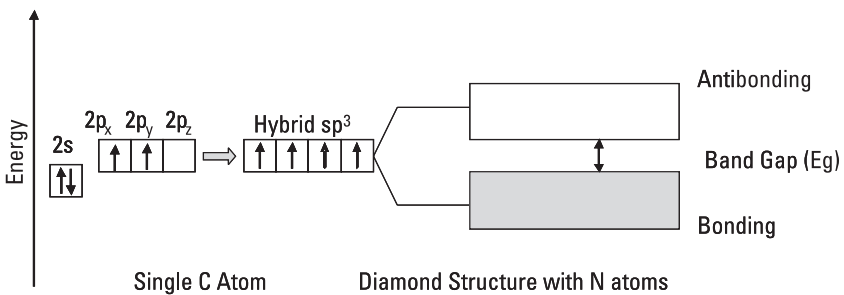


Figure 2.9 Diamond structure showed bonding and antibonding molecular orbitals, leading to a fully occupied valence band and a completely open conduction band, separated by a forbidden gap with energy of E_g . When E_g is large, the material is an insulator; when E_g is small, it is a semiconductor.

valance band, while the empty energy band, called *conduction band*, consists of antibonding molecular orbitals. Between them, there is an energy bandgap (E_g) where no electrons are allowed.

As discussed in the previous section, electrons have to find higher energy states in order to conduct current. Compared with metal, which has a partially filled energy band, this band structure is less favored for conductivity. The valance band is fully filled and there is an energy bandgap before the next available energy level in the conduction band. When an electric field V_0 is applied, an electron cannot move to the next available energy state, which is in the conduction band, unless the potential energy eV_0 is greater than the energy gap. So the electron will have to gain enough energy to find its way to be excited from the valance band to the conduction band before it can carry current. The energy barrier for this excitation is the energy bandgap. When the energy bandgap is large, it is an insulator; when it is small, it is a semiconductor.

2.4 Bonding and Band Structures in Nanocrystal Materials

In previous sections, we discussed atomic bonding in molecules based on LCAO theory and energy band structure in crystalline solids for metals and semiconductors. Molecules consist of a pair to dozens of atoms, whereas any macroscopic crystalline solids usually have atoms on the order of 10^{23} . Nanocrystals, on the other hand, have atoms ranging from several hundreds to tens of thousands, depending on their size. For example, a quantum dot 2 to 3 nm in diameter would have around 1,000 atoms if we assume the lattice parameter is about 2Å to 3Å. If we have a quantum well, a sandwiched thin-film structure, the number of atoms would be much larger because only one dimension is in the nanometer range. Thus, it is clear that the numbers of atoms that we are studying in nanocrystals lies uniquely between molecules and crystalline solids. So does the method that we will use.

Two methods are commonly used to study the bonding and band structure of nanocrystal materials due to their unique sizing properties. One is the bottom-up method, which expands single-molecule calculations such as LCAO theory to nanocrystals. This method requires a lot of computing power and recent developments in supercomputers and software make this approach possible although still expensive. However, even then, we can still only cover about 100 to 1,000 atoms. Otherwise, the matrix is too large to be calculated. The other method is the top-down method, which just adds size-dependent unique properties on top of the standard band structure of macroscopic crystalline materials. This method is very useful for quantum well and other thin-film structures.

2.4.1 Top-Down Method for Quantum Wells and Dots

The top-down method starts with band structure for crystalline solids and adds quantum mechanics to calculate the unique properties due to nano size. Quantum wells and quantum dots are two good examples for demonstrating this method.

Quantum wells are potential wells that confine electron motion to two dimensions in a planar region instead of three dimensions in a free electron case. Quantum wells can be created by placing a semiconductor layer that has a relatively small energy gap between two semiconductor layers with larger energy gaps to form a sandwich structure. The smaller energy gap quantum well is a very thin layer, usually 10\AA to 100\AA thick. This thickness is comparable to the “de Broglie” wavelength of the electrons and holes in the semiconductor and thus the energy levels inside the well are discrete or quantized. The two sandwiched layers should have a bandgap so wide that the electrons and holes confined inside the thin quantum well layer can be treated like the “electron in an infinite 1D potential” problem, which we discussed extensively in the previous chapter.

Based on the band structure of semiconductors as discussed in the previous section, a sketch of the band diagram of quantum wells with different sizes can be constructed, as shown in Figure 2.10. As just discussed, the quantum well can be considered to be a “1D infinite potential well” case. Therefore, the quantized energy inside the well depends on the well size, in this case, the thickness of the thin layer. The thinner the layer is, the more discrete the energy inside the well and the higher the ground state. Therefore, when electrons and holes are combined to form a photon, the thinner quantum well will have high energy and thus frequency.

Example:

Assume the quantum well layer bandgap is 2.5 eV . Calculate the ground state of electrons and holes in the quantum well for thicknesses of 10\AA and 50\AA . The effective mass of electrons (m_e^*) and holes (m_h^*) in this semiconductor is

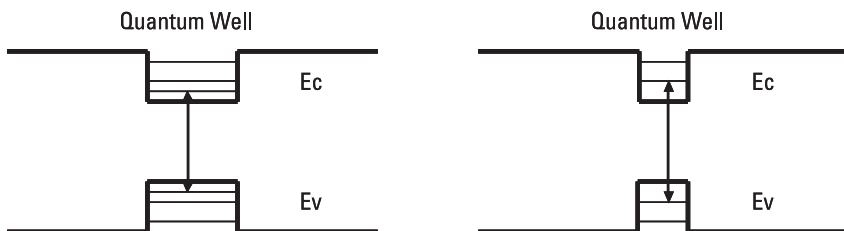


Figure 2.10 Schematic drawing of energy bandgap for quantum well of different sizes.

0.1 and $0.5m_e$. If electrons and holes (both in ground state) are combined, what is the energy and wavelength of the photons emitted from the quantum well?

1. A 10\AA quantum well can be considered to be a 1D infinite well with size of 10\AA . The ground state of the electron is

$$E_1^e = \frac{n^2 h^2}{8m_e^* a^2} = \frac{1 \times (6.63 \times 10^{-34})^2}{8 \times 0.1 \times 9.1 \times 10^{-31} \times (10 \times 10^{-10})^2}$$

$$= 6.038 \times 10^{-19} \text{ J} = 3.77 \text{ eV}$$

The ground state of the hole is

$$E_1^h = \frac{n^2 h^2}{8m_e^* a^2} = \frac{1 \times (6.63 \times 10^{-34})^2}{8 \times 0.5 \times 9.1 \times 10^{-31} \times (10 \times 10^{-10})^2}$$

$$= 1.208 \times 10^{-19} \text{ J} = 0.75 \text{ eV}$$

The energy of photons emitted from the 10\AA quantum well is

$$E_p = \Delta E = 2.5 + 3.75 + 0.75 = 7 \text{ eV}$$

The wavelength of photons emitted from the 10\AA quantum well is

$$\lambda = \frac{Ch}{E} = \frac{3 \times 10^8 \times 6.63 \times 10^{-34}}{7 \times 1.6 \times 10^{-19}} = 1.776 \times 10^{-7} \text{ m} = 177.6 \text{ nm}$$

2. A 50\AA quantum well can be considered to be a 1D infinite well with size of 50\AA . The ground state of the electron is

$$E_1^e = \frac{n^2 h^2}{8m_e^* a^2} = \frac{1 \times (6.63 \times 10^{-34})^2}{8 \times 0.1 \times 9.1 \times 10^{-31} \times (50 \times 10^{-10})^2}$$

$$= 2.415 \times 10^{-20} \text{ J} = 0.15 \text{ eV}$$

The ground state of the hole is

$$E_1^h = \frac{n^2 h^2}{8m_e^* a^2} = \frac{1 \times (6.63 \times 10^{-34})^2}{8 \times 0.5 \times 9.1 \times 10^{-31} \times (50 \times 10^{-10})^2}$$

$$= 0.483 \times 10^{-20} \text{ J} = 0.03 \text{ eV}$$

The energy of photons emitted from the 50Å quantum well is

$$E_p = \Delta E = 2.5 + 0.15 + 0.03 = 2.68 \text{ eV}$$

The wavelength of photons emitted from the 50Å quantum well is

$$\lambda = \frac{Ch}{E} = \frac{3 \times 10^8 \times 6.63 \times 10^{-34}}{2.68 \times 1.6 \times 10^{-19}} = 4.639 \times 10^{-7} \text{ m} = 463.9 \text{ nm}$$

From this example, it is clear that when the quantum well size is small, the energy quantization inside the well is very significant, leading to photons with very different energy being emitted from this semiconductor (the normal bandgap is 2.5 eV versus 7.0 eV for the quantum well). Note also that the photon energy emitted from the quantum well is heavily impacted by the quantum well layer thickness with a $1/a^2$ relationship. However, when the layer size increases, the delta energy compared with a regular semiconductor bandgap decreases. When the layer thickness is much larger (say, 500Å) than de Broglie wavelength of electrons and holes, the energy quantization effect almost disappears. In this case, the quantum well behaves just like a normal semiconductor just as Bohr's correspondence principle predicted.

A quantum dot, on the other hand, is a nanocrystal embedded either in a wide bandgap semiconductor material or insulator. Therefore, it can be considered to be a 3D infinite potential well. The size effect on quantum dots is even more significant than that on quantum wells because quantum dots were limited in all three dimensions. The ground state energy is actually three times higher than that of a quantum well and the quantumization of energy is also more significant. Therefore, quantum dots have been widely used for light emitting diodes (LEDs), lasers, solar cells, optical devices, and displays. Because a quantum dot makes use of more discrete or quantized energy than a quantum well, the light emitted by a quantum dot usually has a narrower spectrum and thus is presented as clear color and sharper image. Furthermore, as the quantum size gets smaller, the ground state energy in the valence and conduction bands gets higher; thus, the light emitted when they are combined will have higher energy and thus smaller wavelength. Figure 2.11 shows that with the bulk bandgap of a 1.8-eV semiconductor, it can only absorb and emit the red light spectrum. However, when we have quantum dots of nanocrystals ranging from 30 to 200 nm, the entire visible light spectrum can be covered. The energy bandgap increases from 1.8 eV on bulk materials to almost 4 eV at the 30-nm size of quantum dots, which leads to a wide range of light-emitting and -absorbing spectra. If desired, readers may use this as another example to calculate the chart in Figure 2.11.

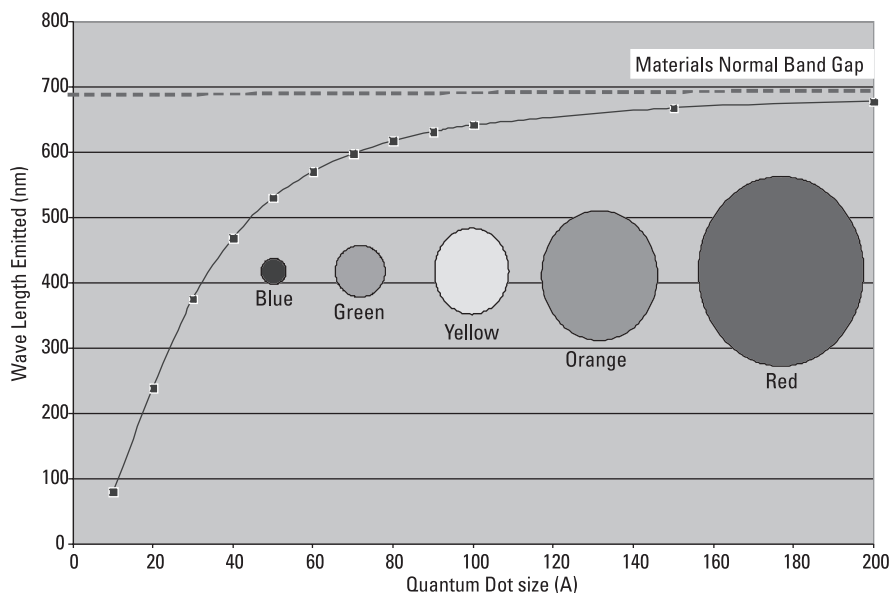


Figure 2.11 Energy bandgap increase with decrease of quantum well thickness, indicating that the quantum effect is significant at small feature size.

2.4.2 Bottom-Up Method for Carbon-Based Nanocrystals

The bottom-up approach uses the same LCAO theory that we used when we studied molecules, but extends it to a much larger number of atoms. Recent developments in computer power and software as well as better computational and approximation methods have made the calculations achievable in a reasonable amount of time. This method is more frequently used on nanocrystal structures, such as carbon nanotubes and the buckminsterfullerene molecule. To study this better, let's start with carbon-based molecular structures in different dimensions.

2.4.2.1 Diamond: 3D Carbon-Based Structure

Diamond can be considered a 3D carbon-based structure. Just like the Si we discussed in the preceding section, carbon atoms form sp^3 hybrid orbitals and thus a covalent bonding with its four neighbors in a tetrahedral structure. Because of its strong covalent bonding, diamond is very hard (high bulk modulus) and not easily deformed (covalent bonding is directional). In addition, because of the wide bandgap of the diamond band structure, it is not only a good electric insulator but also crystal clear (no visible light can be absorbed due to the wide bandgap). A more detailed discussion can be found in the preceding section.

2.4.2.2 Graphite: 2D Carbon-Based Structure

Although it is also carbon based, graphite has very distinct properties from diamond. In graphite, carbon atoms choose sp^2 hybridization instead of the sp^3 hybridization in diamond. Therefore, each carbon atom will form an in-plane covalent bonding with its three neighbors (120 degrees with each other) and thus a hexagonal network within the plane.

The extra unpaired electron left (remember sp^2 hybridization only uses three out of the four valence electrons each carbon atom has) will form an interlayer π bonding. These π bonding molecular orbitals and π^* antibonding molecular orbitals are barely touching each other according to theoretical calculations, so graphite is essentially a gapless semiconductor. In reality, however, π bonding and π^* antibonding molecular orbitals are not only touching each other, but also have about a 40-mV overlap, which makes graphite a good conductor in the direction perpendicular to the plane due to overlapping conduction and valance bands. This interlayer π bonding is also very weak and easily deformed, which leads to high anisotropic mechanical properties. For example, graphite is a good lubricant if we apply shear stress along the plane.

2.4.2.3 Carbon Nanotube: 1D Carbon-Based Nanostructures

Since their discovery in the 1990s, carbon nanotubes (CNTs) have drawn much interest both from scientific incentives and practical applications [2]. It is probably the most famous nanostructure with many unique size-dependent properties widely discussed, including mechanical, electrical, and thermal properties.

A carbon nanotube, in reality, is a rolled-up form of one-atomic-layer-thick graphite, called a 2D graphene sheet. As shown in Figure 2.12, depending on how it is wrapped up, we can classify a CNT using a pair of indexes called chiral vector (n, m) . Shown in Figure 2.12, chiral vector \mathbf{C}_h is the vector that represents the circumference of the nanotube defined as

$$\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2$$

where \mathbf{a}_1 and \mathbf{a}_2 are the unit vectors of the hexagonal lattice. When $m = 0$, the CNT configuration is called zigzag, whereas when $n = m$, it is called armchair. Any other arbitrary n and m combination is just referred to as chiral.

Conceptually, it is not difficult to image that the chiral vector \mathbf{C}_h or the way the carbon nanotube has been wrapped up from the graphene sheet will impact the out-of-plane π bonding and thus its band structure. Both theoretical calculation and experimental results have proven that carbon nanotubes can be metallically conductive if n minus m is a multiple of 3. Otherwise, it will be a semiconductor. Therefore, all armchair configurations with $n = m$ are metallic. It has also been discovered that band structure is also a function of nanotube

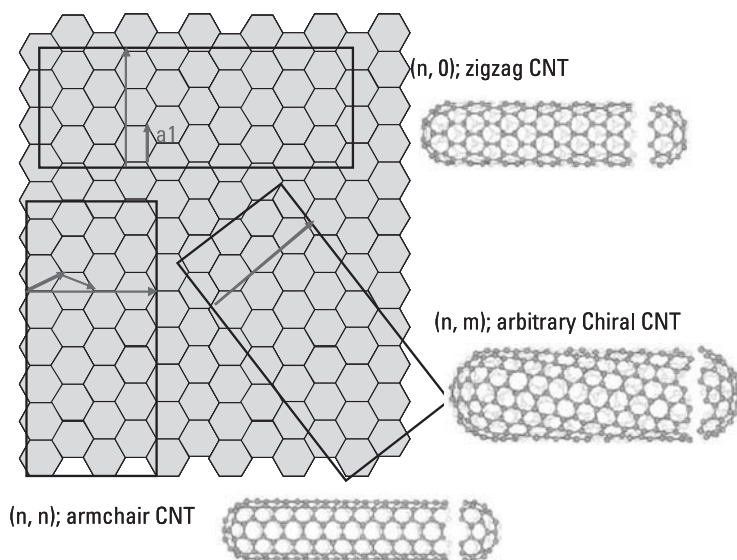


Figure 2.12 Schematic illustration of carbon nanotubes with different chiral angles and how they are achieved through rolling a graphene sheet in different directions.

radius. Some chiral tubes with a semiconductor configuration will turn into metallic with a tube at a very small radius, possibly due to interactions with curvature for π bonding. The detailed calculations for CNT band structures are presented in Chapter 8.

2.4.2.4 C_{60} Fullerene Ball: 0D Carbon-Based Nanostructure

Almost identical to a mini soccer ball, C_{60} fullerene consists of 20 hexagons and 12 pentagons. As shown in Figure 2.13, each pentagon is neighboring five hexagons, and each hexagon borders three pentagons and three hexagons in an alternative way. Therefore, there are totally two different kinds of bonding, one bordering pentagons and hexagons and one between two hexagons. Both types of bonds are intermediate between C–C single bonds and C=C double bonds, indicating a sp^2 hybridization. The molecular orbitals and their energy can be calculated based on LCAO/HMO theory. Using the same methodology we used in the previous section but now with 60 atoms, we can get similar results: 60 atoms will provide 60 molecular orbitals, among which 30 orbitals are above α and 30 orbitals below α . Each atom will also just provide one unpaired electron and thus a total of 60 electrons will occupy 30 molecular orbitals. Therefore, all molecular orbitals below α are fully occupied, while the top 30 orbitals are empty—a perfect example for HOMO and LUMO. Ironically, this calculation was done even before the C_{60} fullerene ball was discovered experimentally. Later, an *ab initio* calculation based on density functional theory

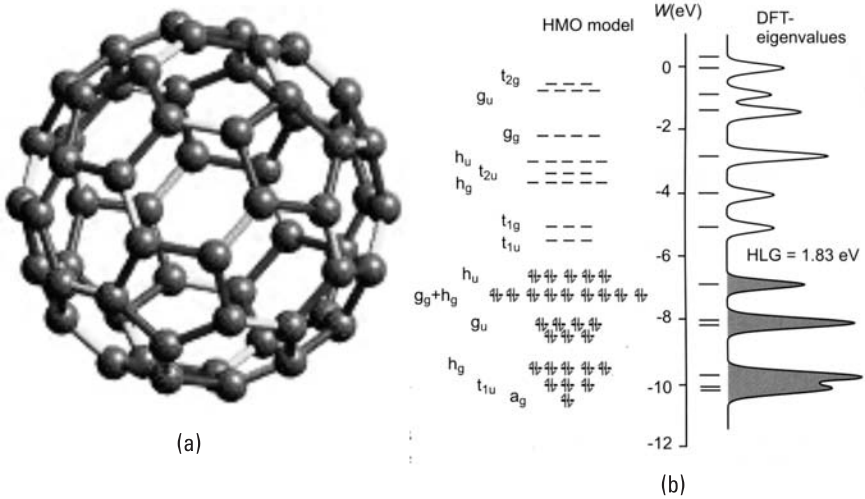


Figure 2.13 (a) Schematic of C₆₀ fullerene ball and (b) its energy level calculated based on LCAO/HMO theory (left) and computational modeling based on density function theory (right). (Reprinted with permission from [1]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

demonstrated that the results from LCAO/HMO are very good approximations as shown in Figure 2.13. More details can be found in [1].

References

- [1] Waser, R. (Ed.), *Nanoelectronics and Information Technology*, New York: Wiley-VCH Verlag GmbH & Co. KGaA, 2005, p. 137.
- [2] Iijima, S., "Helical Microtubules of Graphitic Carbon," *Nature*, Vol. 354, 1991, pp. 56–58.

3

Surface Science for Nanomaterials

In previous chapters, we discussed the atomic bonding and band structures that lead to many size-dependent properties for nanocrystalline materials. The other important perspective for nanocrystalline materials is their surface and surface chemistry due to their smaller size. In this chapter, we first briefly introduce crystal structure and crystallography since most readers should already have learned this during introductory materials science courses. Then we focus on the difference between bulk and surface, the origin of surface energy, and the chemistry and thermodynamics associated with it.

3.1 Crystal Structure and Crystallography

To understand the difference between bulk and surface, we need to discuss crystal structure and crystallography first. Most readers have already learned about this topic in the introductory courses for materials science, so we only go over some basics very briefly.

3.1.1 Crystal Structures

If we consider an atom or a group of atoms as a point mathematically, all crystalline materials can then be considered as a repeating pattern of points in the space called a *lattice*. Groups of lattices can be classified into seven crystal systems and 14 Bravais lattices, which represent the translational symmetry of a crystal. The properties of bulk materials are primarily determined by their crystal structures.

Among the 14 Bravais lattices, most pure metals have one of three lattice structures: face-centered cubic (FCC), body-centered cubic (BCC), or hexagonal close packed (HCP). As shown in Figure 3.1, both FCC and BCC structures have a cubic unit cell with a length a called a *lattice constant*. FCC structure has atoms occupying every corner of the cubic structure and the center of each cube face, whereas the BCC structure has atoms at both corners and the cubic center. On the other hand, the HCP structure, shown in Figure 3.1(c), has a sixfold symmetric hexagonal structure with the length a representing an in-plane hexagonal edge and c representing the height between the top and bottom planes. The top and bottom faces of an HCP cell have six atoms at each corner of a hexagon surrounding a single atom in the center. In addition, there is a middle layer in between the top and bottom surfaces with three atoms each sitting in a hollow site among the top and bottom atoms. Thus, each atom has six closest neighbor atoms within the HCP base plane and three atoms each on layers above and below. So in an HCP structure each atom has a total of 12 neighbors. Additional information regarding other types of Bravais lattices can be found in any college-level introductory materials science book for readers who are interested.

3.1.2 Crystallography

To better describe the lattice structure, we have to introduce crystallography—a systematic method for specifying the planes and directions in crystal structures. Let's first define the coordinate system. The basic coordinate system usually consists of three noncoplanar axes along unit cell edges (\mathbf{a} , \mathbf{b} , \mathbf{c}) with origin at a corner of the unit cell. Note that because the unit cell edges may not have the same length and may not be perpendicular to each other, this coordinate system may not be a Cartesian system unless the unit cell is a cubic structure (simple cubic, FCC, or BCC).

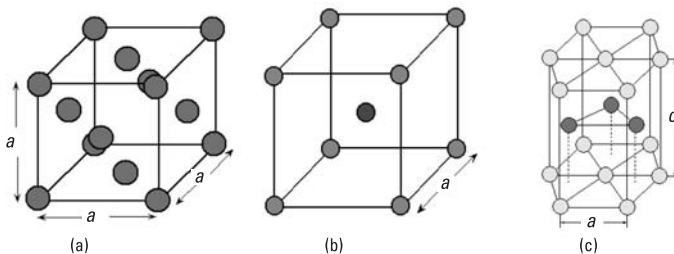


Figure 3.1 Illustration of typical crystal structure: (a) face-centered cubic or FCC structure, (b) body-centered cubic or BCC structure, and (c) hexagonal close-packed or HCP structure.

Once we have defined the coordinate system, it is easy to index any crystallographic direction. We can consider any crystallographic direction as a vector that can be described by a set of integers $[uvw]$ based on the coordinate system as follows:

$$\mathbf{m} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad (3.1)$$

in which \mathbf{a} , \mathbf{b} , and \mathbf{c} are the axes of the unit cell representing the coordinating system; \mathbf{m} is the vector representing that specific crystallographic direction, and u , v , and w are integers, indicating how many steps it is from the origin. Thus, we can index any crystallographic direction using a set of integers $[uvw]$ in the coordinate system of the unit cell. We use $\langle uvw \rangle$ to define a family of crystallographic directions that are identical due to the symmetry of the crystals.

Defining the orientation of crystallographic planes is similar but not as straightforward as indexing crystallographic directions. First, we are using the same coordinate system based on three axes along the unit cell edge (\mathbf{a} , \mathbf{b} , \mathbf{c}). For any given plane, the plane will either parallel or intercept one of the three axes. The reciprocals of interceptions on three axes are used to index the orientation of the crystallographic planes. This set of integers is written in a bracket as (hkl) and is called the Miller index. If the plane is parallel to the axes, the interception is actually infinite and thus the reciprocal number is 0. We also use $\{hkl\}$ to index a family of crystal planes if they are identical due to symmetry of the crystals, just like the crystallographic direction family.

Exercise:

After defining the Miller index for both crystallographic planes and directions, we can now prove that crystallographic direction $[hkl]$ is perpendicular to crystallographic plane (hkl) .

3.1.3 Close-Packed Directions, Planes, and Structures

If we assume atoms are solid balls with the same radius r , we can identify the close-packed direction and planes in all crystal structures.

3.1.3.1 BCC Structures

In BCC structures, because we have atoms occupying all cubic corners and the cubic body center, the only directions in which all atoms (balls) touch each other is cubic body diagonal or $\langle 111 \rangle$. Along any other directions (for example, the $[100]$ cubic edge), atoms do not touch each other. Therefore, $\langle 111 \rangle$ is the close-packed direction. If the atom's radius is r and BCC has lattice parameter a , the relationship between r and a can be easily obtained through the fact that $\langle 111 \rangle$ is the close-packed direction. We have

$$\sqrt{3}a = 4r \Rightarrow a = \frac{4}{\sqrt{3}} r \quad (3.2)$$

From this equation, it is clear that $a > 2r$ and it confirms again that atoms are not close packed along cube edge $\langle 100 \rangle$ with length a .

Further, we can calculate volume packing density (also called atomic packing factor or APF) for BCC structures and each plane. For BCC structures, we have eight atoms occupying each cubic corner that is shared by eight neighbor cubes and one atom at the center. Therefore, the total number of atoms in a BCC structure is $8 \times (1/8) + 1 = 2$ atoms. So the APF for a BCC structure is volume of 2 atoms/volume of cube, which is

$$\frac{2 \times \left(\frac{4}{3} \pi r^3 \right)}{a^3} = \frac{2 \times \frac{4}{3} \pi r^3}{\left(\frac{4}{\sqrt{3}} \right)^3 r^3} = 0.68$$

In a similar way we can calculate the packing density of a close-packed plane. For example, in plane $\{110\}$, we have two atoms in terms of total occupied areas. Therefore, the packing density is

$$\frac{2 \times (\pi r^2)}{a \times \sqrt{2}a} = \frac{2 \times \pi r^2}{\left(\frac{4}{\sqrt{3}} \right)^2 r^2 \sqrt{2}} = 0.833$$

while for plane $\{100\}$, we only have one atom so the total packing density is

$$\frac{1 \times (\pi r^2)}{a^2} = \frac{\pi r^2}{\left(\frac{4}{\sqrt{3}} \right)^2 r^2} = 0.589$$

Hence, $\{110\}$ is more close packed than $\{100\}$. In fact, $\{110\}$ is the plane that has the highest packing density in BCC structures and thus is the close-packed plane.

In summary, in a BCC structure, the close-packed direction is $\langle 111 \rangle$, the close-packed plane is $\{110\}$, and each atom has eight closest neighboring atoms.

3.1.3.2 FCC Structures

Atoms in FCC structures occupy all cubic corners and all face centers. Therefore, the atoms are touching each other along face diagonals or the crystallographic directions of $\langle 110 \rangle$. The relationship between an atom's radius r and lattice parameter a can be expressed as follows:

$$\sqrt{2}a = 4r \Rightarrow a = 2\sqrt{2}r \quad (3.3)$$

Based on this relationship, it is obvious that atoms are not touching each other along other directions including cube edge $\langle 100 \rangle$; thus, $\langle 110 \rangle$ is the close-packed direction.

Using the same method, we can calculate the atomic packing factor for FCC structures. Again, we have eight atoms occupying each cubic corner that is shared by eight neighbor cubes and six atoms at the face center shared by only two cubes. Therefore, there are a total of $8 \times (1/8) + 6 \times (1/2) = 4$ atoms in FCC structures. The FCC packing density is, therefore,

$$\frac{4 \times \left(\frac{4}{3} \pi r^3 \right)}{a^3} = \frac{4 \times \frac{4}{3} \pi r^3}{(2\sqrt{2})^3 r^3} = 0.74$$

For a close-packed plane, it is not as straightforward as for a BCC configuration. For example, we can easily calculate the packing density for the (100) plane. In this plane, we have two atoms so the packing density is

$$\frac{2 \times (\pi r^2)}{a^2} = \frac{2 \times \pi r^2}{(2\sqrt{2})^2 r^2} = 0.7854$$

The packing density of (110) can be calculated as 0.555 using the same method. The (111) plane also has two atoms but it is a triangle. So the density is

$$\frac{2 \times (\pi r^2)}{\frac{1}{2} \times (\sqrt{2}a)^2 \times \frac{\sqrt{3}}{2}} = \frac{4 \times \pi r^2}{(2\sqrt{2})^2 r^2 \times \sqrt{3}} = 0.907$$

Hence, for FCC structures, the close-packed plane is $\{111\}$ and the close-packed direction is $\langle 110 \rangle$.

3.1.3.3 HCP Structure

The HCP structure was described in the previous section. Because atoms are touching each other on the base plane, it is clear that the close-packed direction is along all edge directions as well as diagonal in the base plane. Therefore, the edge length of the base plane is equal to two times the atom's radius: $a = 2r$. Furthermore, if we need the middle plane touching both top and bottom base planes, the height of hexagonal structure c will have to satisfy the following:

$$c = 2 \times \sqrt{a^2 - \left(\frac{2}{3} \times \frac{\sqrt{3}}{2} a\right)^2} = \frac{2\sqrt{6}}{3} a = \frac{4\sqrt{6}}{3} r \quad (3.4)$$

The close-packed plane is the base plane. For a hexagon, it contains $6 \times (1/3) + 1 = 3$ atoms. So its packing density can be calculated as

$$\frac{3 \times (\pi r^2)}{3a^2 \times \frac{\sqrt{3}}{2}} = \frac{2 \times \pi r^2}{4r^2 \times \sqrt{3}} = 0.907$$

For HCP volume packing density, each hexagonal cell contains 3 atoms in the middle plane; 2 atoms in the top and bottom center, which is shared by two adjacent cells; and 12 atoms at the corner, each shared with 6 cells. So the total number of atoms in each cell is $3 + (2 \times 1/2) + (12 \times 1/6) = 6$. Thus, the APF for HCP structures is

$$\frac{6 \times \left(\frac{4}{3} \pi r^3\right)}{3a^2 \times \frac{\sqrt{3}}{2} \times c} = \frac{8\pi r^3}{3\sqrt{2}a^3} = \frac{8\pi r^3}{3\sqrt{2}(2r)^3} = 0.74$$

3.1.3.4 Close-Packed Structures

Interestingly we cannot help noticing that the volume packing density of HCP structures is identical to that of FCC structures. So is the packing density for a close-packed plane. We already know that in the HCP structure, each atom has 12 closest neighbor atoms—6 in the same base plane and 3 each at the hollow site of both the top and bottom planes as shown in Figure 3.1(c). It can also be mathematically proven as well as illustrated graphically that 12 is the maximum number one atom can have for closest neighbors and, thus, HCP is a close-packed structure.

Now if the FCC structure has the same packing density as an HCP structure, then the FCC must also be a close-packed structure. If we plot the

close-packed plane (111), we can clearly see that any atom in this plane has six closest neighbor atoms touching each other. It also has three atoms each on the hollow site on the crystal planes above and below. The only difference between the FCC and HCP structures is that neighboring atoms on the crystal planes above and below are not in the same set of hollow sites in the FCC structure. In other words, the top and bottom atoms are not aligned together as in the HCP case. If we call the HCP structure a stacking structure of close-packed plane (the base plane) with a sequence of ABABAB . . . , the FCC structure is nothing but a similar stacking structure of close-packed plane (111) with a sequence of ABCABC Therefore, the FCC structure is also a close-packed structure with a different stacking sequence.

In contrast, the BCC structure has a packing density of only 0.68 with only eight closest neighboring atoms. Therefore, the BCC structure is not a close-packed structure. The close-packed direct, plane, packing density and its closest neighboring atoms for the BCC, FCC, and HCP structures are summarized in Table 3.1.

3.2 Surface Crystallography

We have discussed crystal structure and crystallography in bulk materials. Now let's take a close look at what happens on the surface.

3.2.1 Surface Structure for Close-Packed Structures

As discussed in the previous section, both FCC and HCP are close-packed structures in which each atom has 12 closed neighbors. If the surface is their close-packed planes, that is, {0001} (Note: the indexing of crystallographic planes and directions in HCP structures are different with standard cubic structures. This is why we have 4 digit planes. Interested readers can find it in any materials science textbook.) for HCP and {111} for FCC as shown in Figure

Table 3.1
Summary of Close-Packed Planes and Directions, APF, and Closest Neighbors for BCC, FCC, and HCP Structures

	BCC	FCC	HCP
Close-packed directions	$\langle 111 \rangle$	$\langle 110 \rangle$	Hexagonal edge
Close-packed planes	{110}	{111}	Base plane {0001}
Atomic packing factor (APF)	0.68	0.74	0.74
Number of closest neighbors	8	12	12

3.2, each surface atom will be identical and have a coordination number (CN) of nine, or nine closest neighbors—six of them in the plane and three atoms in the very next plane underneath the surface. In addition, the atoms in the very next layer underneath the surface will have a CN of 12, which means that it is already like atoms in the bulk. So this layer in theory can already be considered as being a bulk layer.

If the surface plane is the $\{100\}$ plane in FCC structures, each atom has eight closest neighboring atoms with four on the same plane and four underneath. Therefore, it is apparent that coordinate numbers of surface atoms are different depending on the crystallographic planes.

3.2.2 Surface Structure for BCC Structures

Now, let's take a look at BCC structures. In a BCC structure, a surface atom on the $\{100\}$ plane will only have four closet neighbors in the plane underneath the surface. There are no atoms on the surface that are the closest neighbors. So the CN of an atom on a surface plane of $\{100\}$ is 4 as shown in Figure 3.3(a). The atoms in the next layer, however, have CN of 8 as bulk material.

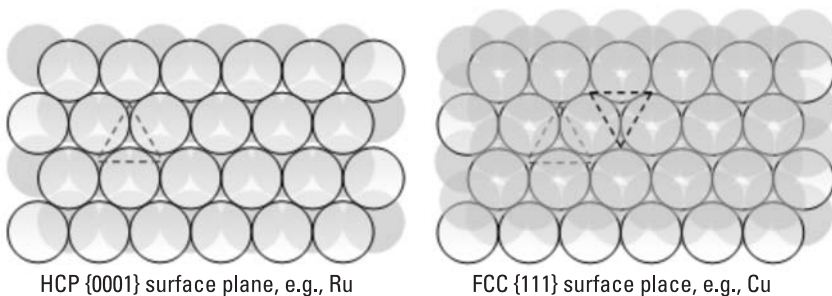


Figure 3.2 Illustration of crystallographically close-packed surface on HCP and FCC. Different shades indicate different atomic layers.

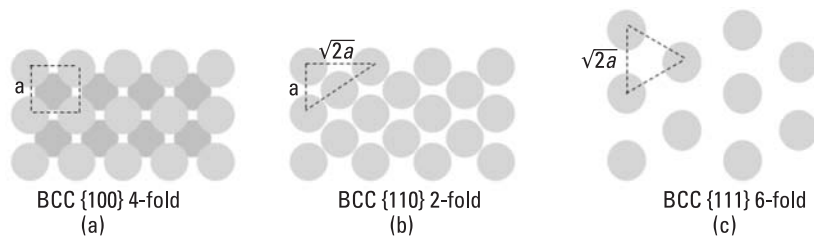


Figure 3.3 Illustration of crystallographic surface planes for the BCC structure: (a) $\{100\}$ plane with fourfold, (b) $\{110\}$ plane with twofold, and (c) $\{111\}$ plane with sixfold symmetry.

If the surface is one of the $\{110\}$ planes in the BCC structure as shown in Figure 3.3(b), the CN of surface atoms will be 6, with four closest neighbors in the surface plane and two closest neighbors on the layer underneath the surface. For the $\{111\}$ surface planes, each atom also has four closest neighbors, although it is less obvious. [Figure 3.3(c)]. (Readers interested in pursuing this can research it for themselves.) Therefore, the atoms in the (110) surface plane have more coordinating numbers than (111) and (100) , just as in the case of the FCC structure. This turns out to be a very important phenomenon in surface crystallography, which we discuss later.

3.2.3 Surface Symmetry

The unit cell on the surface can be described using two vectors, \mathbf{a}_1 and \mathbf{a}_2 . Using FCC as an example, as shown in Figure 3.4, the unit cell for the $\{100\}$ plane is a square with fourfold symmetry, whereas the $\{110\}$ plane has a unit cell of rectangular shape with twofold symmetry. The close-packed plane of $\{111\}$ has the highest symmetry of sixfold.

The surface unit cell and its symmetry will be also one of the determining factors for surface energy, especially when there is another layer (whether this is done through absorption or deposition) on top. We discuss this later.

3.3 Surface Energy

The origin of the surface energy arises from the fact that surface atoms are situated in a different environment compared with their bulk counterparts. For example, any atom in bulk materials with FCC or HCP structure will have 12 nearest neighbors and, thus, 12 interatomic bonding. However, for the atoms on the surface, as discussed previously, they will have fewer neighbors due to

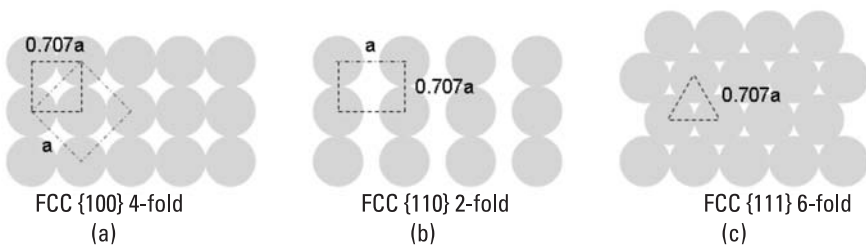


Figure 3.4 Illustration of crystallographic surface planes for the FCC structure: (a) $\{100\}$ plane with fourfold, (b) $\{110\}$ plane with twofold, and (c) $\{111\}$ plane with sixfold symmetry. Compared with BCC structures, the symmetries are same but the unit cell parameters and packing density are different.

their unique terminating locations. As a result, those atoms will have some unsaturated or dangling bonds, which, in turn, will add some extra energy to the surface atoms compared with those in the bulk materials. This extra energy is the origin of the surface energy.

Every finite object has surface area and thus carries surface energy. However, the portion of this “extra” surface energy in overall system energy is not the same.

Exercise:

Assume we have made some perfect spherical particles with a uniform radius R . So each individual particle has a volume of $\frac{4}{3}\pi R^3$ and surface area of $4\pi R^2$. If we have two types of particles with different sizes, 1 μm and 1 nm, which one should consider the surface energy more significantly?

Giving the same volume of V , the total surface area can be described as

$$\frac{V}{\left(\frac{4}{3}\pi R^3\right)} \times (4\pi R^2) = \frac{3V}{R}$$

indicating that for a fixed total volume of a spherical particle, the total surface area and thus total surface energy are proportional to $1/R$. So if we have two types of particles with different sizes, 1 μm and 1 nm, the total surface energy of 1-nm particles is 1,000 times that of 1- μm particles!

This example demonstrates that surface area and energy associated with surface are extremely significant when the material’s size being studied is on the nanometer scale.

For convenience, the surface energy (ϕ) is usually defined as this extra energy per unit area. Therefore, the total energy associated with the surface can be described as $E = \phi * S$, where S is surface area. Because surface energy is an “extra” energy, nature will always want to decrease this term and thus the overall energy. This unfavorable contribution can be minimized through one of the two ways or a combination of both:

1. Minimizing the surface energy per unit surface area (ϕ) through either exposing predominant surface planes that have a lower surface energy or altering the local surface atomic geometry to reduce the surface energy (surface reconfiguration);
2. Reducing the amount of surface area exposed (S).

We discuss these methods in the next two sections together with surface-related physics, chemistry, thermodynamics, and materials science, especially in

the nanomaterials regime. The remainder of this section discusses the preferred crystallographic planes for minimizing the surface energy, and Section 3.4 explores different ways to alter the surface geometry for reconfiguration. Section 3.5 introduces ways to reduce the exposed surface area thermodynamically.

3.3.1 Crystallographically Preferred Surface

We found in the previous section that when the surface plane is a crystallographically close-packed plane, it usually has the highest coordination numbers or closest neighboring atoms, which means the fewest dangling/unsaturated bonds. This can also be considered in a slightly different perspective. We all know that the close-packed plane has the highest number of atoms and thus the highest packing density in the plane. Therefore, they usually also have the highest interatomic layer distance, which indicates relatively weak layer-to-layer bonding. However, this does not necessarily mean that the close-packed plane has a minimal surface energy per unit area since it has more atoms per unit area due to the high packing density.

Using the FCC structure as an example again, we can compare the relationship between the number of dangling bonds, interatomic layer spacing, and an estimated surface energy per unit area. As summarized in Table 3.2, although the close-packed plane has the least number of dangling bonds per atom, it also has the most number of atoms per unit area. According to first principle estimation, it may not necessarily be the crystallographically preferred surface plane with minimum surface energy per unit area.

3.3.2 Wulff Constructions and Equilibrium Shape for Nanoparticles

The surface energy of any arbitrary surface plane can be estimated using Wulff constructions. To simplify the problem, we can consider a simple cubic structure

Table 3.2
Summary of Surface Properties of Different Crystal Planes in FCC Structures
(a is the lattice constant and ϵ is the bonding energy)

	(111)	(100)	(110)
Number of dangling bonds	3	4	5
Interatomic layer spacing	$0.577a$	$0.5a$	$0.354a$
Surface energy per atom	$3*0.5\epsilon$	$4*0.5\epsilon$	$5*0.5\epsilon$
Number of atoms per unit area	$2 \text{ atoms}/(0.866a^2)$	$2 \text{ atoms}/a^2$	$1 \text{ atom}/(0.747a^2)$
Surface energy per unit area	$3.464 (\epsilon/a^2)$	$4 (\epsilon/a^2)$	$3.347 (\epsilon/a^2)$

with a lattice parameter of a . As shown in Figure 3.5, any arbitrary surface plane that is at an angle θ to the close-packed plane will have excessive broken bonds compared with those on the close-packed plane, primarily due to the broken bonds at the steps. For each atom (point A) at close-packed plane (100), there is one broken bond. However, any atoms at a step (point B) will have one extra broken bond. If we assume the unit length on the arbitrary plane is at an angle θ with close-packed plane, then the total broken bond per unit length is $\cos \theta/a + \sin \theta/a$. Therefore, the surface energy per unit area is $\phi = (\cos \theta + \sin \theta)/a^2 \cdot (0.5\epsilon)$. This can be plotted as ϕ versus θ . Clearly the surface energy per unit area ϕ is at a minimum when $\theta = 0$, indicating that the minimum surface energy is achieved at the close-packed plane orientation and any deviation from this surface plane ($\theta \neq 0$) will increase the surface energy. Similarly we can apply this method to any low index planes in any crystal structure—they should all be at the minimal surface energy cusps due to the relatively close-packed structure.

Another way to represent the relationship between surface energy and any arbitrary crystallographic plane is to plot the surface energy ϕ as a 3D contour in the crystal structures. This is called a Wulff construction. For example, in Figure 3.6, we can plot FCC [100] as the x axis and [011] as the y axis. Therefore, the [111] axis will be between the x and y axes. Along each direction, we can plot the surface energy of that crystallographic plane normal to this direction as shown in Figure 3.6. The Wulff construction is a type of polar representation of surface energy ϕ and has been widely used to predict the equilibrium shape of single-crystal particles, especially nanoparticles.

Assume a single crystal particle has several crystallographic surface planes S_1, S_2, \dots , with surface energy of ϕ_1, ϕ_2, \dots , the equilibrium shape of the particle will have to minimize the total surface energy given by $E = \sum \phi_i S_i$. As shown in Figure 3.6(a), for any arbitrary point i in the ϕ plot, we can connect

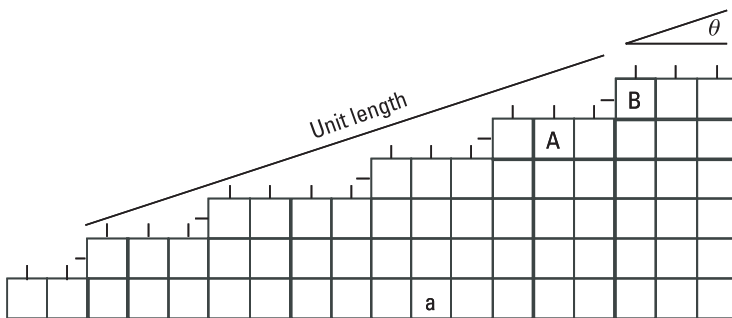


Figure 3.5 Illustration of excessive surface energy for a surface deviated from a low index plane.

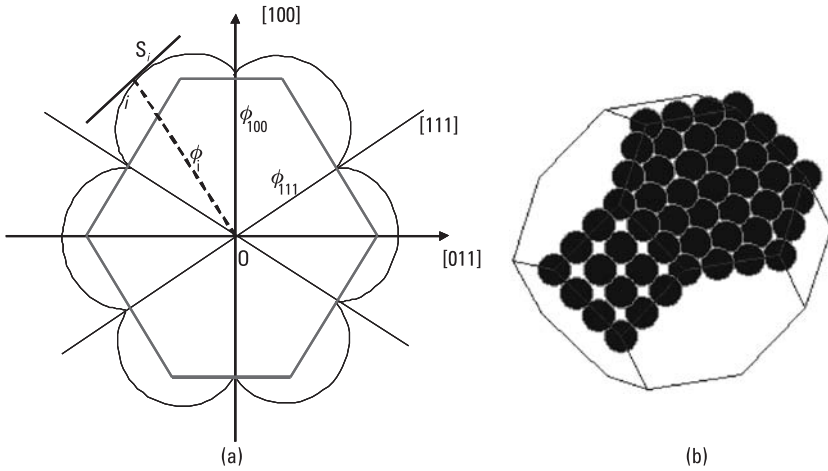


Figure 3.6 (a) Wulff constructions of ϕ plot to obtain the equilibrium shape of nanoparticles in 2D and (b) equilibrium shape for nanoparticles showing $\{111\}$ -type surfaces and six smaller, $\{100\}$ -type surfaces.

origin O and point i and draw a plane S_i that is normal to the radius Oi . Here the length of Oi is the surface energy ϕ_i and the surface area is the plane S_i . Thus, to minimize the overall surface energy term $E = \sum \phi_i S_i$ and to obtain the equilibrium shape for the particle with crystallographically preferred surface planes, we need to draw a normal plane through every point on the ϕ plot and find the most inner envelope among all planes. As discussed earlier, the ϕ plot usually has cusps at the low index close-packed plane—this will inevitably lead to large facets of those planes in the equilibrium shape as shown in Figure 3.6(b).

Because surface energy plays an important role, nanoparticles with equilibrium shape often consist of the crystallographically preferred surface planes. For example, in FCC nanoparticles, the $\{111\}$ and $\{100\}$ facets are always observed, but not the $\{110\}$ surface planes, as shown in Figure 3.7 [1], which is in a very good agreement with the Wulff construction (Figure 3.6). Furthermore, when the particle size is even smaller, the edge energy induced by atoms located on the edge sites between two adjacent facets becomes more significant, leading to another extra term in total energy and equilibrium shape deviated from the Wulff construction.

3.4 Surface Reconfigurations

In the previous section, we used a broken bond model to introduce surface energy. In a real materials system, surface energy is much more complicated.

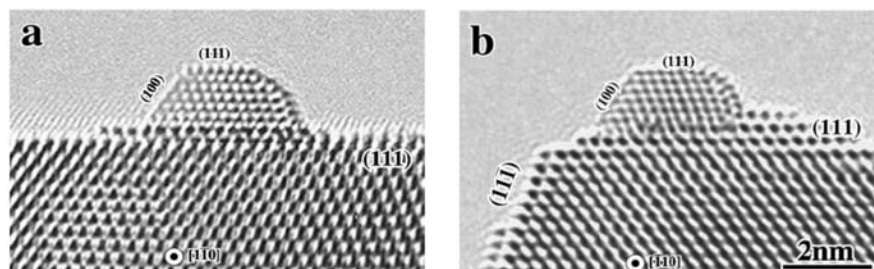


Figure 3.7 (a, b) Nano FCC metallic particles (Au) showing (111) and (100) surface. (Reprinted with permission from [1]. Copyright 2005 Springer.)

Surface relaxation and reconfiguration will dictate the surface symmetry, crystallography, and thus its energy. In general, there are two types of surface reconfigurations. One is homogeneous relaxation and reconstruction, meaning that surface atoms assume different positions, crystal structures, and symmetry compared to the bulk counterpart to relax the surface energy. The other is heterogeneous adsorption through a second medium on the surface. We discuss both methods in this section.

3.4.1 Surface Relaxation and Reconstructions

In a perfect crystal, the equilibrium position of an atom, as discussed in Section 3.1, is determined by minimizing the total energy or where the external forces from its neighboring atoms balance each other. Therefore, it is intuitive that the surface atoms will not be in the same exact equilibrium positions as their bulk ones since the external force acting on the surface atoms is different due to the terminating position. We usually call it *surface relaxation* when atoms in the entire surface layer shift either vertically or laterally relative to the layer underneath, while their relative position within the surface layer remains unchanged. On the other hand, if there is a surface structure or symmetry change in addition to a position shift, we usually refer to it as a *surface reconstruction*.

3.4.1.1 Vertical and Lateral Surface Relaxation

There are two different kinds of relaxation. Relaxation along the normal direction (vertical to the surface) is easy to understand because all of the external forces come from one direction. As shown in Figure 3.8(a), the first surface layer will have smaller interatomic layer spacings. Because now the first layer is closer to the second layer, the force between the first and the second layer is different from the force between the second and the third, pushing the second layer also close to the third layer. Therefore, the interatomic layer spacing between the second and the third layer is also different from the ideal bulk crystal. However,

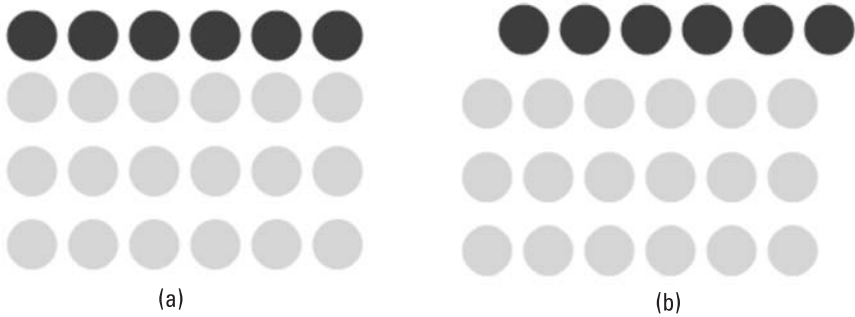


Figure 3.8 Illustration of (a) vertical and (b) lateral relaxed surface planes showing every atom in the entire surface plane shift to the layer underneath while maintaining the relative position within the surface plane.

the delta of the spacing compared with bulk is gradually decreasing until, after three to five layers, it is completely diminished. Therefore, the surface layer, or to be exact, the surface layers are not just limited to the very top layer, but should also include the first three to five atomic layers because they all have different interatomic layer spacing compared with the bulk due to relaxation normal to the surface direction.

Similar things can happen with lateral relaxation, although the change is less apparent. Surface atoms now not only change their positions in the direction normal to the surface, but also move away from their equilibrium positions laterally within the surface plane to further minimize the dangling bond and relax the energy as shown in Figure 3.8(b). However, in both cases, the relative atomic position and thus surface structure and symmetry in the surface plane remain unchanged after relaxation.

3.4.1.2 Surface Reconstructions

Under some circumstances, an atom might even change its surface structure and symmetry to further reduce the surface energy, as shown in Figure 3.9. Unlike relaxation, the phenomenon of reconstruction involves a change in the periodicity of the surface structure.

One example is silicon. With a diamond-like FCC structure, silicon usually cleaves along the (111) and (100) planes as per Wulff constructions. If the surface plane is (100), as shown in Figure 3.10, the as-cleaved surface without lateral relaxation has a 2×2 square unit cell, and each atom has two dangling bonds from diamond structures. After relaxation, two neighboring Si atoms reunite with two dangling bonds to form a new bond. Hence, the number of dangling bonds is decreased by a factor of 2, leading to a much lower surface energy.

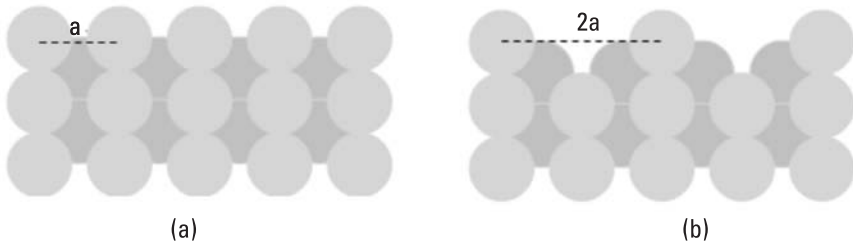


Figure 3.9 Illustration of surface structure (a) before and (b) after surface reconstruction.

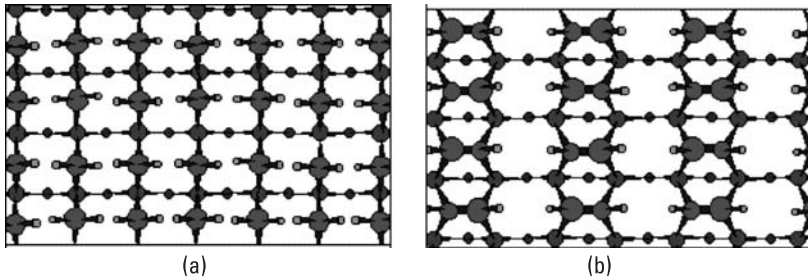


Figure 3.10 Silicon (100) surface when in an (a) unrelaxed and (b) relaxed situation. The different sizes of circles indicate the relative location to the top surface. (Reprinted with permission from University of Minnesota Physics Department website, http://groups.physics.umn.edu/stmlab/gallery/surface/surface_fr.html.)

Due to the formation of the new bond, two Si atoms now form a pair. They are closer to each other than the column of silicon atoms on the left or the right side of this pair. Hence, the symmetry of the 2×2 square unit cell is no longer held. Instead, the lower energy surface plane has reconstructed itself into 2×1 symmetry with full and empty columns of atoms alternating with each other, as shown in Figure 3.10(b). Therefore, surface reconstruction has not only changed the positions of the surface atoms but also the symmetry of surface unit cell. Readers can find more details in [2].

3.4.1.3 Wood's Notation for Surface Reconstruction

Surface reconstructions change not only the positions of surface atoms but also the structure and symmetry of the surface. If the basic vectors of the original unit cell in bulk are \mathbf{a}_1 , \mathbf{a}_2 before reconstruction, and \mathbf{b}_1 , \mathbf{b}_2 are the basic vectors of the new unit cell after surface reconstruction, the relationship between $(\mathbf{a}_1, \mathbf{a}_2)$ and $(\mathbf{b}_1, \mathbf{b}_2)$ can be described in matrix notation such that

$$\begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \times \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} \quad \text{or} \quad \begin{cases} \mathbf{b}_1 = m_{11}\mathbf{a}_1 + m_{12}\mathbf{a}_2 \\ \mathbf{b}_2 = m_{21}\mathbf{a}_1 + m_{22}\mathbf{a}_2 \end{cases} \quad (3.5)$$

If the angle between two basic vectors before and after the reconstructions remains unchanged, meaning that the original and the new unit cell have the same or closely related symmetry, (3.5) can have a simple linear solution, which is called *Wood's notation*. Figure 3.11(a) shows the surface unit cell that is just $2a$ of the bulk one. Thus, Wood's notation is $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Figure 3.11(b) shows a rotated unit cell after reconstruction so the matrix is $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

Wood's notation can be used to describe overlayer symmetry with either homogeneous (surface constructions) or even heterogeneous (adsorption) surface layers. We discuss the absorption case in Section 3.4.2.

3.4.1.4 Surface Layers and Nanoparticles

We point out here that surface reconstruction can also impact atomic layers underneath. As one can imagine, when the surface atoms move away from their original equilibrium position, their bonding with the atoms on the layers below will inevitable be affected. Similar to normal relaxation, this effect will gradually disappear after three to five layers. Therefore, when we discuss surface layers, we will include all of the layers that have different bondings/spacings from the bulk and thus different physical and chemical properties compared to the bulk materials. This usually encompasses about three to five atomic layers after surface relaxation.

Exercise:

Estimate the percentage of surface versus bulk atoms in a nanometer-sized particle.

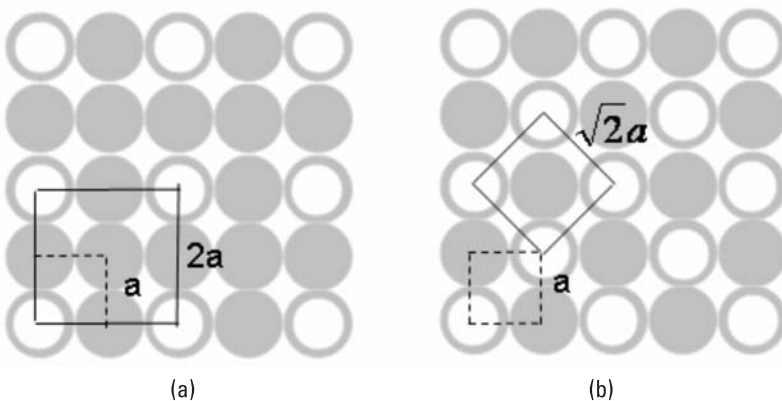


Figure 3.11 Illustration of surface unit cell and Wood's notation after the surface reconstruction (a) with enlargement and (b) with rotation.

To simplify the problem, let's assume a simple cubic structure with cubic equilibrium shape as predicted by Wulff constructions. If we assume the interatomic spacing is 2.5\AA and the surface region is four layers, we will have a 10\AA or 1-nm-thick surface layer. If we have a 10-nm cubic-shaped nanoparticle, the total number of atoms is $40 \times 40 \times 40$ or 64,000 atoms in this particle. However, because we have four layers of surface atoms on each side, the atoms in bulk regions are only $32 \times 32 \times 32$ or 32,768. So in total we have 51% bulk and 49% surface atoms.

We can calculate and plot the percentage of surface atoms versus particle size as shown in Figure 3.12. When particle size is large (in the micrometer regime), the surface atoms account for less than 1% of overall atoms. So the properties of the particles are generally dominated by bulk. But as the particle size decreases to the nanometer regime, the percentage of surface atoms increases dramatically. As we just calculated, almost 50% of the atoms are on the surface in a 10-nm particle. When the particle size is down to 2 nm, all of the atoms in the particle will have to be considered to be surface atoms. Therefore, the properties of nanoparticles will be completely dominated by the surface. This again demonstrates that surface energy and properties are significant factors when the material dimension is in the nanometer range. Ironically, people usually define nanotechnology as dimensions less than 10 nm. From our estimation here,

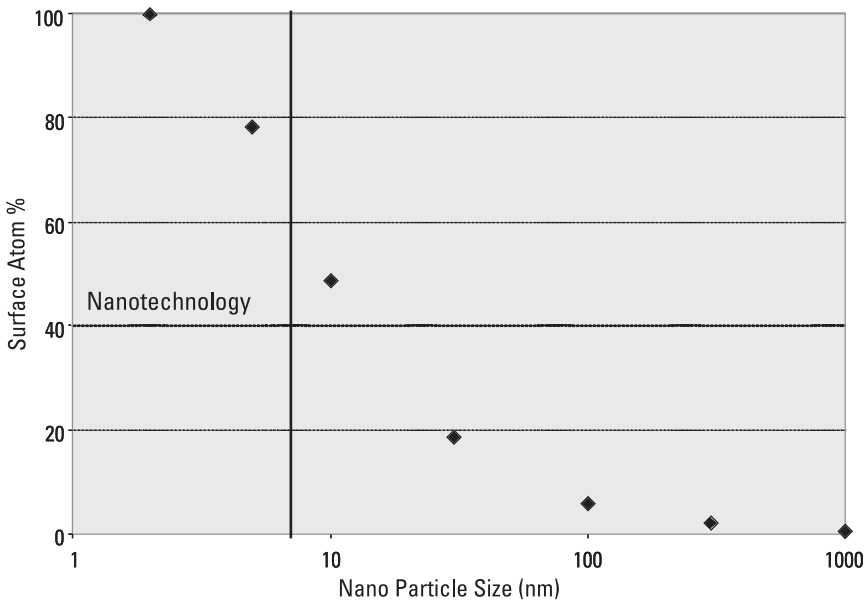


Figure 3.12 The percentage of surface atoms increases dramatically as the particle size decreases.

it appears that 10 nm happens to be the transition point where surface properties start to play a dominant role. (Surface atoms are over 50%.) From this point of view, nanotechnology can also be considered surface technology.

3.4.2 Adsorption

Surface relaxation and reconstruction leads to a homogeneous reconfiguration that often occurs on an atomically clean surface in vacuum with no other phases involved. In reality, however, there are always other phases such as air, vapor, and so forth. Therefore, another method for decreasing surface energy per unit area is *adsorption*, a process in which atoms or molecules from other phases become adsorbed on the surface. As we know, unlike their counterpart in bulk, the atoms on the surface are not completely surrounded by their closest neighbors and, thus, can attract other atoms or molecules and decrease surface energy through adsorption. The adsorption layers can also have their own unit cell and surface symmetry if adsorbates form an ordered structure in the preferred adsorption sites. This structure can be described by Wood's notation. So this can be considered to be heterogeneous surface reconstructions through adsorbate atoms with ordered structures.

3.4.2.1 Adsorption Sites

A surface offers different types of sites where adsorbate atoms can potentially land. As shown in Figure 3.13, the first type is the on-top site, which is directly on the top of surface atoms. In this case, the adsorbates will form a single bond with just one atom directly below. The second type is called the bridging site between two surface atoms, at which point adsorbate will form bonds with both atoms. The third is the hollow site between three or four atoms to form more

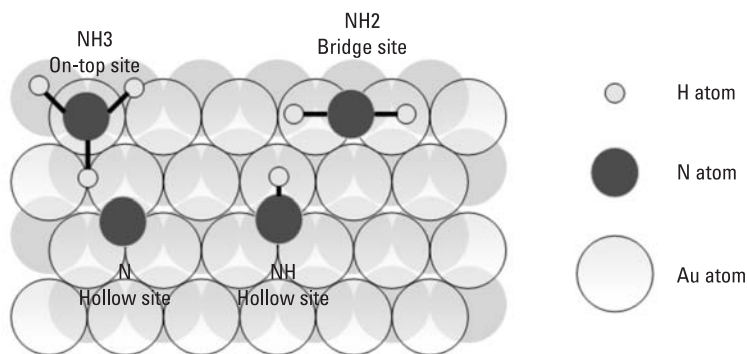
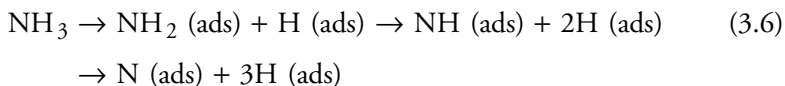


Figure 3.13 Illustration of the adsorption of NH_3 on an FCC metal surface. Depending on the disassociation of NH_3 , the adsorbates will take different adsorption sites.

bonds. The level of disassociation of hydrogen atoms in NH_3 , the adsorbate can occupy different surface sites and form different bonds with substrate atoms.

Figure 3.13 shows the adsorption of NH_3 on an FCC Au surface. The nitrogen atom in the NH_3 molecule has a pair of $2s$ electrons that are not used to form bonding with hydrogen atoms. Therefore, when adsorbed at a metallic surface, it can share this pair of electrons with one surface atom at the on-top site through a weak bond. The NH_3 molecule can also form one of the following adsorbates by disassociating hydrogen atoms at different levels:



As the number of hydrogen atoms bonded to the nitrogen atom is reduced, the adsorbed species will tend to move into a higher coordination site on the surface, such as the bridge site and hollow site, and thereby form more and stronger covalent bonding with surface atoms.

3.4.2.2 Physical and Chemical Adsorption

Now we know that surface atoms can attract adsorbates through interatomic bonding. Depending on the nature of the bonding, the process of adsorption can be divided into physical adsorption or physisorption (usually van der Waals bonding) and chemical adsorption or chemisorption (typically covalent bonding).

Table 3.3 summarizes the difference between physical and chemical adsorptions due to their bonding nature. Both are very important to surface chemistry and nanosurface science. For example, ALD (atomic layer deposition) in semiconductor processing uses physical adsorption of precursors on substrate, whereas SAM (self-aligned monolayer) utilizes chemical adsorption for nanofabrications. Both will be discussed in detail in the chapters on nanomaterials fabrication (see Part II). Note also that adsorption has a very close relationship with wetting, especially on the nanometer scale. Interested readers can find more details in [3].

3.5 Surface Area and Surface Thermodynamics

In the previous two sections, we discussed methods for decreasing surface energy per unit area through either exposing crystallographically preferred surface or surface reconfiguration (relaxation or reconstructions), both homogeneously or heterogeneously. In this section, we discuss surface energy minimization through exposing less surface area. This can be accomplished also through either homogeneous or heterogeneous means. Although the homogeneous method involves a

Table 3.3
Comparison of Chemical and Physical Adsorption

	Chemisorption	Physisorption
Temperature range (over which adsorption occurs)	Virtually unlimited (but a given molecule may effectively adsorb only over a small range).	Near or below the condensation point of the gas (for example, Xe < 100 K, CO ₂ < 200 K).
Adsorption enthalpy	Wide range (related to the chemical bond strength); typically 40 to 800 kJ mol ⁻¹ .	Related to factors such as molecular mass and polarity but typically 5 to 40 kJ mol ⁻¹ .
Crystallographic specificity (variation between different surface planes of the same crystal)	Noticeable variation between crystal planes.	Virtually independent of surface atomic geometry.
Nature of adsorption	May be irreversible.	Reversible.
Saturation uptake	Limited to one monolayer.	Multilayer uptake possible.
Kinetics of adsorption	Varies—often an activated process.	Fast, since it is a nonactivated process.

shape/configuration change to minimize the surface area exposed, the heterogeneous method introduces a second phase on top of the surface such as wetting. The first method is especially important for nanomaterials where surface energy is dominant. The latter involves many classical surface thermodynamics and kinetics.

3.5.1 Surface Area in Nanomaterials

As mentioned in the previous section, surface energy is a very important facet of total energy. It becomes even more significant when the scale of materials we discussed is down to the nanometer size. As shown in Section 3.4, for a particle with a size of 3 to 5 nm, its properties will be dominated by the surface because most of the atoms need to be considered as surface atoms. Therefore, for nanomaterials, nature is always trying to minimize the overall energy by reducing the surface area exposed. For example, nanoparticles are very easy to coalesce together as a group to minimize the total surface area exposed.

Another interesting example is the carbon nanotube (CNT). As introduced briefly in Chapter 2, the CNT can be considered a rolled-up version of a single sheet of graphite called graphene. As we know, graphite has sp^2 hybridization so it can form three single bonds (σ type) with three in-plane neighbors. The

other unpaired electron will form a weak π bond between layers. When we have a single sheet of graphene, this weak bonding has been broken, leading to a high surface energy. In addition, there are also edge atoms that are in termination positions in two dimensions. So these atoms have even more dangling bonds and thus higher energy. On the other hand, when we roll this graphene sheet into a CNT, all of the edge atoms will again be connected through σ -type covalent bonds with the atoms along the other side. Furthermore, once the single sheet is rolled up as a tube, those unpaired electrons will interact with each other inside the tube and form π bonds to reduce the surface energy. In addition, most of the CNT will have a half buckeye ball (C_{60}) on the tip of the tube. This is also to minimize the existence of dangling bonds of edge atoms. The surface energy can be further reduced if a multiwall carbon nanotube (MWCNT) is formed instead of single wall (SWCNT) because less surface area is exposed as numerous cylinders are tightly stacked together.

Surface energy also plays an important role for thin-film deposition. The addatom will take positions not only just on the crystallographically preferred plane to decrease the surface energy per unit area, but also first take the preferred site such as steps and kinks to expose less surface area and thus minimize the overall surface energy ($E = \phi^*S$). This will be discussed in detail in the section on nanomaterials fabrication (Part II).

In general, the overall energy, including the surface energy, will have to be minimized—resulting in what is called a *thermodynamic equilibrium state*. Here we assume our readers already have some basic knowledge of thermodynamics, such that we focus on some surface-related phenomena.

3.5.2 Nanoparticle Nucleation

This is a classical case of thermodynamics that involves surface energy. It has been described in great detail in many college-level introductory thermodynamics textbooks. Here we will explain a bit about what this means to nanoparticles.

As shown in Figure 3.14(a), a nanoparticle has nucleated from the parent phase. The overall Gibbs free energy change can be described as

$$\Delta G = -V_{np} \Delta G_v + S_{np}^* \phi \quad (3.7)$$

where V_{np} and S_{np} are the volume and surface area of nanoparticles. If we assume the nanoparticle is a spherical shape with a radius of r , we can plot the ΔG as a function of particle size r as shown in Figure 3.14(b). Because the volume of the nanoparticle is proportional to r^3 while the surface area is proportional to r^2 , the surface energy is dominating in the region of the small radius, whereas the volume of the Gibbs energy will become significant when radius r is large enough. Therefore, there exists a critical radius of r^* , below

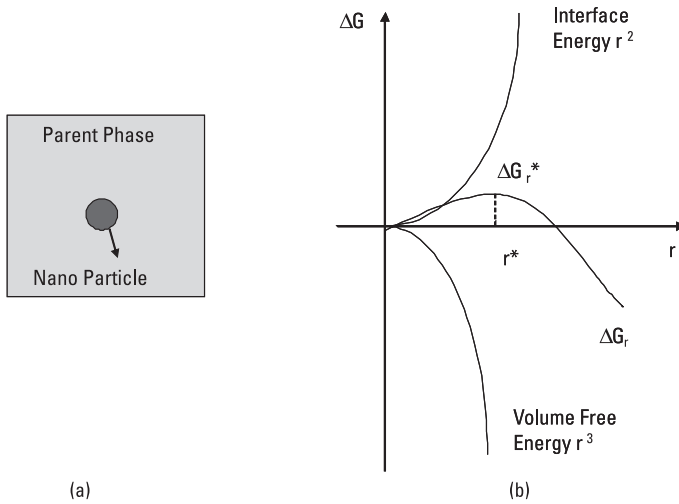


Figure 3.14 (a) Illustration of a nanoparticle nucleated from the parent phase and (b) the relationship between Gibbs free energy and the particle radius, indicating that when the radius is below r^* , the nanoparticle is thermodynamically unstable.

which overall ΔG is positive, indicating that a particle size smaller than r^* is thermodynamically unstable and will be dissolved back into the parent phase. Thus, r^* is the smallest particle size we can achieve thermodynamically. It can be easily expressed as

$$r^* = \frac{2\phi}{\Delta G_v} \quad (3.8)$$

Hence, if we want to push the critical size of nanoparticle r^* even smaller, we have to either decrease the surface energy per unit area for nanoparticle ($\phi \downarrow$) or increase the delta of Gibbs free energy per unit volume ($\Delta G_v \uparrow$). Both of these methods have been used in fabrication of nanoparticle-size materials—we will provide details in the following chapters.

3.5.3 Wetting

Wetting is another surface phenomenon that is not only used in nanomaterial fabrications, but also in many other cases in real life ranging from frying pans to car waxes to glues and solders. Whether one phase can “wet” on the substrate surface thermodynamically is determined completely by the surface energy of these two phases as well as the interface energy between these two phases. It is a combination of decreasing surface area exposed and surface energy per unit area to minimize the overall energy.

3.5.3.1 Young's Equation and Contact Angles

Using a liquid drop on a solid surface as an example (see Figure 3.15), the contact angle or wetting angle is a characteristic value to evaluate how well the liquid (or phase II) spreads on the substrate surface in an ambient vapor environment.

From the balance of surface tension, we can easily get

$$\cos \theta = \frac{\phi_{SV} - \phi_{SL}}{\phi_{LV}} \quad (3.9)$$

where ϕ_{SV} and ϕ_{LV} are surface energy of the substrate and liquid drop, and ϕ_{SL} is the interface energy between the substrate and liquid drop. This is called Young's equation.

3.5.3.2 Complete and Partial Wetting

In Young's equation, if we have $\phi_{SL} + \phi_{LV} = \phi_{SV}$, we will have contact angle $\theta = 0$ degree. In this case, the liquid or second phase can completely spread on the substrate, called complete wetting. This is the thermodynamic equilibrium state. Moreover, if the surface energy of the solid is higher than the sum of the surface energy of the liquid plus the interface energy between solid and liquid, the system would rather to have two interfaces ($S-L$, $L-V$) instead of $S-V$ to minimize surface energy. This is the ideal case for painting, coating, soldering, and depositing films in nanomaterial fabrications. To promote wetting, one can do one of the following:

1. Raise ϕ_{SV} by cleaning the surface. For example, flux in soldering removed oxide from the surface or cleaning the substrate interface before growing or depositing thin films.
2. Lower ϕ_{SL} —one example is to add Sn in solder to form intermetallic compound.
3. Lower ϕ_{LV} by adding surfactant to the lower surface energy of the second phase.

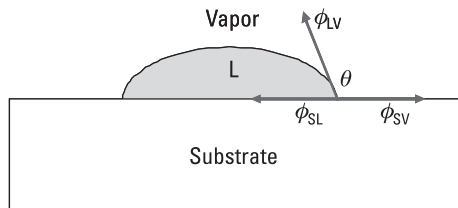


Figure 3.15 Illustration of contact angle and Young's equation.

On the other hand, if we have $\phi_{SV} + \phi_{LV} < \phi_{SL}$, the surface of the second phase and substrate is preferred compared with the S - L interface. Thus, the liquid or second phase cannot spread on the substrate at all, called complete nonwetting. The most popular complete nonwetting case is the nonstick frying pan. To promote dewetting, we need similar measures to either raise ϕ_{SL} or lower ϕ_{SV} or ϕ_{LV} .

One famous example for wetting and dewetting is hydrophobic and hydrophilic interfaces, or in other words, water-loving and water-hating interfaces. Materials with a hydrophilic surface exhibit an affinity for water. The surface chemistry allows these materials to absorb water and be wetted, forming a water film. Hydrophilic materials also possess a high surface tension and have the ability to form “hydrogen bonds” with water. On the other hand, hydrophobic surfaces have little tendency to adsorb water and water tends to “bead” on the surfaces. Hydrophobic materials possess low surface tension values and lack active groups in their surface chemistry for formation of “hydrogen-bonds” with water.

References

- [1] Akita, T., et al. “TEM Observation of Gold Nanoparticles Deposited on Cerium Oxide,” *J. of Materials Science*, Vol. 40, 2005, pp. 3101–3106.
- [2] Dabrowski, J., and Mussig, H.-J., *Silicon Surfaces and Formation of Interfaces*, Singapore: World Scientific Publishing, 2000, p. 135.
- [3] Luo, J., and Chiang, Y.-M., “Wetting and Prewetting on Ceramic Surfaces,” *Annual Review of Materials Research*, Vol. 38, 2008, 227–249.

4

Nanomaterials Characterization

We have discussed the fact that many objects have particle-wave duality, such as photons and electrons. Human beings have long been using those objects to make direct or indirect observations in material characterization. For example, for thousands of years people have used visual inspection (photons from light) to check the color of quenched steel to determine its mechanical properties. Later, metallographic technology was introduced to study the microstructure of materials through optical microscopy (photon, but with optical lenses). Modern materials science has developed many more advanced techniques to study materials structure down to the molecular and even atomic level. Electron microscopes use electron beams to explore materials microstructures, and XPS and X-ray diffraction are used to study atomic and lattice information by means of X-ray photons. These techniques are especially useful for nanomaterials science because all of these beams have smaller wavelengths and thus higher resolution compared to visible light, which enables us to explore fine, structural details right down to the nanometer scale.

In this chapter, we first introduce some fundamental principles of materials characterization such as diffraction. Then we explain the unique concerns and considerations for nanomaterials characterization and the related new techniques, such as X-ray diffraction electron microscopes, and surface analysis technologies such as SIMS, Auger, and XPS.

4.1 X-Ray Diffraction for Nanomaterials Characterization

Diffraction, one of the key characteristics of wave nature in particle-wave duality, is essentially due to the interaction between two or more waves that have the

same frequency and a certain relationship in phase angles. Diffraction has been demonstrated for both light and electron beams through a double split in Young's experiment. In this section, we discuss X-ray diffraction for crystal structure characterization.

4.1.1 X-Ray Diffraction and the Laue Method

Any 1D wave propagating along the x direction can be described as $A = A_0 \cos(kx - \omega t)$. So the amplitude of the wave follows a cosine curve going through maximum A_0 , 0, and minimum $-A_0$ in one period. Let's examine two waves with the same amplitude A_0 and frequency. If both waves reach the maximum or minimum at the same time, the amplitude will double the original value to reach $2A_0$. This is called *constructive diffraction* and the two waves are "in phase." On the other hand, if two waves happen to be 180 degrees apart, meaning that when one wave reaches the maximum, the other reaches the minimum, the two waves will cancel each other and the net result is zero. This is called *destructive interference* and the two waves are "out of phase."

Diffraction occurs not only in the double split, but also in any periodic structures such as crystal lattices. In 1912, German physicist Max von Laue first examined the crystal structure using X-rays. Based on the observation of the diffraction of visible light, Laue reasoned that if the wave length of an X-ray were similar to that of lattice parameters or interatomic spacing, it should be possible to diffract an X-ray through a crystal lattice. He and his team successfully demonstrated this idea through the experimental setup illustrated in Figure 4.1. A CuS crystal was placed in the path of a traveling white X-ray beam (with variable wavelength) and a photosensitive plate was set up on the other side of the CuS crystal to collect the diffracted X-ray beams. A spot pattern was successfully recorded to prove Laue's hypothesis—not only does an X-ray have wave nature since it can be diffracted, but crystal structures also have periodicity

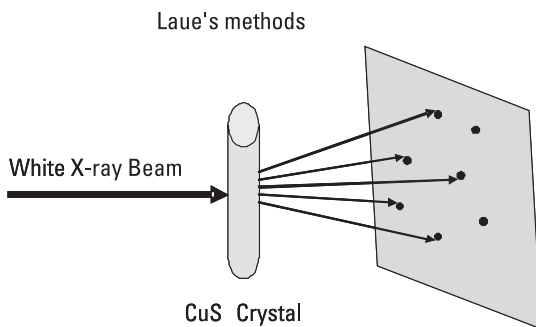


Figure 4.1 Illustration of the Laue method—a white X-ray beam is diffracted by CuS crystals. A photosensitive screen recorded the diffracted spot pattern.

with interatomic spacing similar to an X-ray's wavelength. This method is called the *Laue method*.

Laue's observations stimulated a great deal of interest in the materials science communities because this was the first time X-ray diffraction was demonstrated and fine features inside crystal structures revealed. Shortly after Laue published his results, two British physicists, W. H. Bragg and his son W. L. Bragg, derived the condition for constructive diffraction based on this observation. This is called Bragg's law.

4.1.2 Bragg's Law

As shown in Figure 4.2, the incident beam can be considered to be a planar wave. Beam 1 is reflected at atomic surface plane *A*, whereas beam 2 is reflected at the very next one below—atomic plane *B*. In both cases (beams 1 and 2), the reflection angles are the same as incident angles θ . Because the incident beam is a planar wave, beams 1 and 2 are in phase before they reach the sample. After the reflection, beam 2 has to travel an extra distance represented by the sum of $MO_2 + M'O_2$. From the illustration, clearly this distance $MO_2 + M'O_2 = 2d \sin \theta$. To maintain the in-phase relationship between beams 1 and 2, and thus achieve the constructive diffraction, the extra distance beam 2 travels has

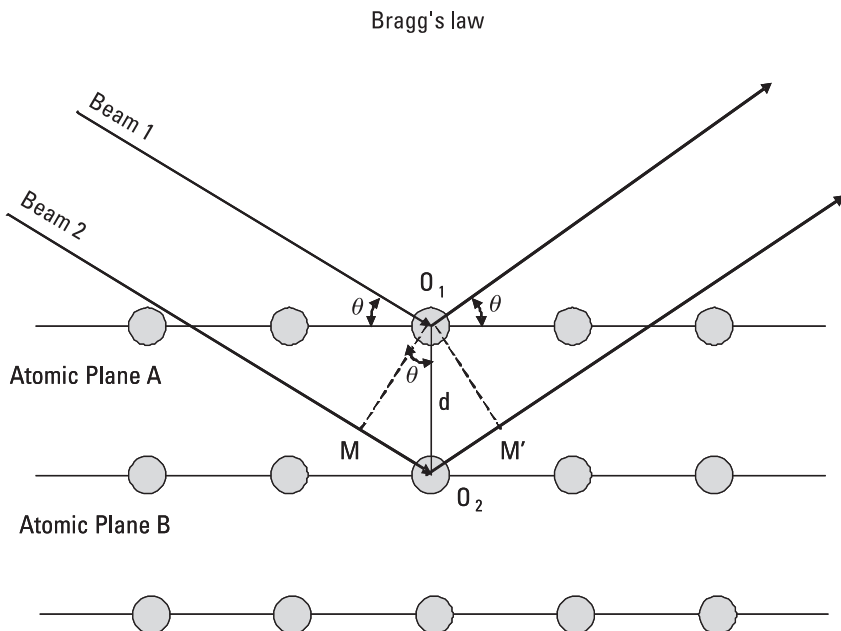


Figure 4.2 Illustration of Bragg's law.

to be an integer number times the wavelength or $n\lambda$. This is called Bragg's law:

$$2d \sin \theta = n\lambda \quad (4.1)$$

where d is the periodicity of structure or in this case, the interatomic spacing; λ is the wavelength of the incident beam; θ is the incident angle; and n is the integer (1, 2, 3, ...). It is apparent that in order to achieve constructive diffraction, the period of structure d will have to be on the same order with wavelength λ . For example, visible light has a wavelength on the order of 5000Å, therefore, the optical microscope can be used down to the submicro level. On the other hand, X-rays are widely used to characterize the periodic lattice structure because their wavelength ($\sim 0.6\text{Å}$ to 0.7Å) is in the same range as lattice parameters (on the order of 2Å to 3Å) in solid materials.

Given an incident X-ray beam with certain wavelength (λ), for a crystal plane that has different interatomic spacing (d), constructive diffraction will only happen at a certain angle of θ . Therefore, if we can detect X-ray intensity at angle θ and we know the incident beam wavelength (λ), we can back-calculate the interatomic layer spacing (d), which will help to reveal the chemical and phase information of the material. This method, derived from Bragg's law, has been widely used in materials characterization, especially for X-ray diffractometry. In this diffraction technique, a known wavelength of monochromatic X-ray radiation is used to determine the unknown spacing of materials. The specimen can be single crystals, polycrystals, or even powder. The intensity of the diffracted beam is recorded at different angles of θ to calculate the spacing and, thus, identify the chemical and phase information of the material being studied.

4.1.3 X-Ray Diffraction in Nanomaterials

Bragg's law derived the conditions for constructive diffraction at angle θ . Now let us take a look at what happens when two beams are not exactly in phase.

As shown in Figure 4.2, the difference of the path traveling between beams 1 and 2 is $2d \sin \theta$. Bragg's law explained that when this delta is the integer number of the wavelength, the two beams are always exactly in phase and constructive diffraction was observed. On the other hand, if the phase difference $MO_2 + M'O_2 = 2d \sin \theta$ is half of the wavelength different or 180° (π) out of phase, the two diffracted beams will completely cancel each other, leading to destructive interference. But what if the delta is a quarter of the wavelength different or 90° ($\pi/2$) out of phase? These two diffracted beams will not cancel each other. So shouldn't we observe some diffracted beam at this angle with a certain intensity? Not really. Although beams 1 and 2 will not cancel each other completely (Figure 4.2), because they are 90° ($\pi/2$) out of phase, beam 3,

which is reflected by the atomic plane C (not labeled in Figure 4.2) that is directly below A and B , will be 180 degrees out of phase with beam 1 so these two beams will cancel each other completely, while beam 2 will be canceled by beam 4, which is reflected by atomic layer D (not shown in Figure 4.2). Thus, all of the beams will still cancel each other and we will observe a complete destructive interference at that angle of θ .

From the preceding discussion, we can draw the following conclusion: For a given set of crystallographic planes, you will observe any diffracted pattern at and only at an angle of θ according to Bragg's law. As soon as the angle deviates from θ , say, to $\theta + \Delta\theta$, one can always find a layer that is $180/\Delta\theta$ layers below so that the reflected beams can cancel each other. Thus, in an ideal case, Bragg's law predicts that mathematically the X-ray diffraction pattern is a standard delta function as shown in Figure 4.3(a). Diffraction can only be observed at the angle of θ for a given set of crystal planes.

The preceding discussion is based on two assumptions. First, this is only true for ideal crystals that have infinite crystal planes and are defect free. As described earlier, when $\Delta\theta$ is small, one will need a large number of atomic layers below the surface to find the right one in order to completely form destructive infringement. Secondly, it also assumes that all reflected beams have the same amplitude regardless of how deep the atomic layers are below the

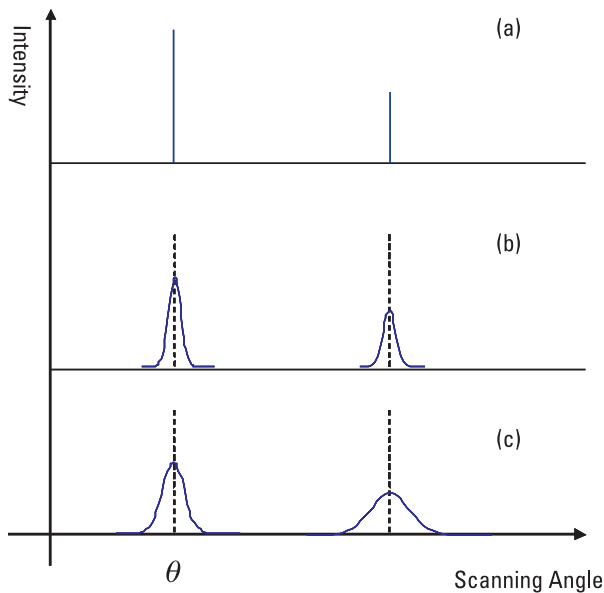


Figure 4.3 X-ray diffraction patterns: (a) linear pattern like the delta function in ideal crystals, (b) wave pack pattern in real crystals, and (c) diffused peak in nanocrystals.

surface. This is only true at the region near the surface, because the incident beam will be scattered inelastically and the amplitude will be dramatically decreased. Thus, in a real case of X-ray diffraction, when $\Delta\theta$ is small, the reflected beam from deep below will have a much smaller amplitude than the one on the surface and, thus, the beams cannot completely cancel each other. As a result, instead of a linear pattern like the delta function, we will observe a wave pack-like diffraction pattern with a peak at the angle of θ but spreading to $\theta \pm \Delta\theta$ as shown in Figure 4.3(b). The total peak width is $2\Delta\theta$.

Things get more interesting when the specimen is a nanocrystal or nanoparticle. First of all, as we discussed in Chapter 3, most of a nanocrystal's atoms are considered to be surface atoms. Secondly, nanoparticles or crystals have small particle or grain sizes and, thus, a limited number of atomic planes. For example, if the interatomic layer distance d is 2.5\AA , nanocrystals with 5 nm only contain about 20 atomic layers. This is far less than needed to form complete destructive interference at the angle near θ . As a consequence, the peak at the angle of θ is greatly broadened and diffused and thus the peak width $2\Delta\theta$ is much larger [Figure 4.3(c)]. In fact, one can estimate the nanoparticle's or nanocrystal's grain size from the peak width. Readers who are interested in this can find more details in [1]. From the first principle estimation, one can imagine that a 5-nm nanocrystal with 20 atomic layers will have a half-peak width $\Delta\theta$ of at least 9° ($180^\circ/20 = 9^\circ$). Therefore, the X-ray peak is spreading across 18° around the angle of θ . This spreading will increase significantly when the particle size is even smaller. Normally people use the *full width at half-maximum* (FWHM) measurement to characterize the diffraction peak and thus estimate the grain or particle size. The extreme case of nanocrystals is glass, which has short-range order and long-range disorder. In this case, the peak is so broadened that all peaks are connected and they can hardly be called peaks anymore—this has been frequently observed in the amorphous X-ray diffraction pattern. Therefore, peak broadening is one of the key characteristics of X-ray diffraction on nanocrystals. From the width of a peak or more specifically the FWHM, one can obtain information about nanocrystal grain size or nanoparticle size. This method has been widely used in nanomaterial characterization.

Note, however, that grain size is not the only cause for peak widening. Other factors, such as micro strains, will also cause peak widening. But this discussion is beyond the scope of this book. Interested readers can find more details in many X-ray diffraction textbooks such as [1].

4.2 Electron Microscopy for Nanomaterials Characterization

As discussed in the previous section, the resolution is on the same order of the wavelength of the incident beam. For example, since the wavelength of visible

light is around 2000\AA to 8000\AA , one can observe objects down to the micron or submicron level using an optical microscope with a magnification of about 1,000 times. Beyond that, the image resolution is greatly reduced. On the hand, according to quantum mechanics, electrons have a particle-wave duality nature with the wavelength inversely related to electron momentum. Therefore, one can imagine that if we accelerate electrons under high voltage, an electron beam should have a smaller wavelength to provide us high resolution at the nanometer scale. This is the fundamental principle of electron microscopy. There are several different kinds of electron microscopes, such as the transmission electron microscope, scanning electron microscope, and scanning tunneling electron microscope. In this section, we first provide a brief overview of the interaction between electron beams and solids. Then we quickly review these three kinds of electron microscopes and their applications to nanomaterials characterization. Interested readers can find more details about microscopes in [2].

4.2.1 Interaction Between Electron Beams and Solids

When a high-energy electron incident beam shines on a solid specimen, the electrons will interact with the specimen and generate many signals that can reveal information about the materials. According to the properties and directions, these signals can be divided into the categories shown in Figure 4.4 and discussed next.

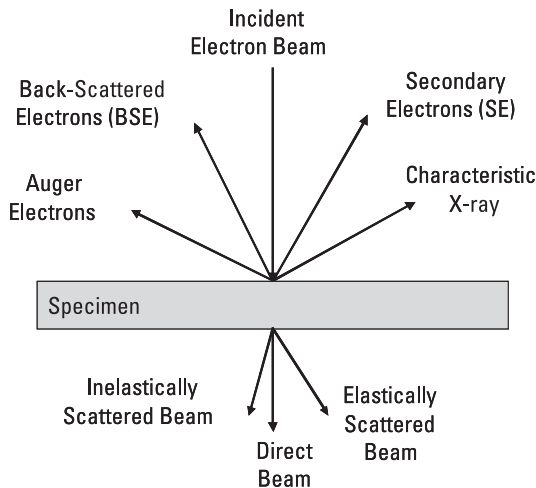


Figure 4.4 Illustration of interaction of an incident electron beam with a specimen to generate different kinds of signals that can be used for materials characterization.

4.2.1.1 Transmitted Electrons

Some electrons will pass through the specimen if the sample is thin enough and the electrons have high enough energy. The transmitted electrons contain much information on the material's microstructure. For example, electrons may be diffracted by a lattice and thus can form a diffraction pattern according to Bragg's law. Also, electrons can also be scattered either elastically or inelastically as they pass through the specimen. Therefore, the energy loss can also be used to reveal the chemical composition of materials. Transmitted electrons mainly are used in transmission electron microscopy for imaging, diffraction pattern and electron energy loss spectroscopy.

4.2.1.2 Backscattered Electrons

If the sample is too thick for transmission, some of the electrons that have entered the specimen may get absorbed, whereas others can be scattered inside the sample and eventually escape. Those electrons are called backscattered electrons and can be collected and analyzed for their directions and energy loss in order to characterize the crystallographic orientation and the chemical composition of the material's surface.

4.2.1.3 Secondary Electrons

When the incident electrons impact the sample surface with sufficient energy, it will ionize the atoms and excite additional electrons from outer orbits. Those electrons are called secondary electrons. They are the main means used in scanning electron microscopy to study surface morphology.

4.2.1.4 X-Rays

While secondary electrons are the electrons excited from the outer shell by the incident electron beam, it is also possible to ionize the electrons from the inner shells. However, the atoms with missing inner shell electrons are usually unstable. The electrons in the outer shell will instantly fill the inner shell to minimize the total energy. The energy delta between the outer shell and inner shell is released through X-ray photons. Because this energy delta is characteristic to atoms, the X-ray spectrum can be used to determine the material's chemical composition and its quantitative amount in the impact zone near the surface.

4.2.1.5 Auger Electrons

Auger electrons are very similar to X-rays. The only difference is that instead of X-ray photons, the energy released further excites additional electrons, called Auger electrons. Therefore, this process involves a three-electron process. First, an incident energy beam excites an electron from the inner shell. Second, an outer shell electron instantly hops into the inner shell and releases energy as the delta between the energy level of the inner and outer shells. Third, if this

released energy is large enough, it will excite another outer shell electron. This is why the Auger effect is a three-electron process that involves both interstate and intrastate transitions. Auger electrons can be used to characterize the chemical composition of a few atomic layers right below the surface.

4.2.2 Transmission Electron Microscope (TEM)

Shortly after the discovery of the particle-wave duality of electrons, people realized that the wave characteristics of electrons could be used to build a microscope that could surpass the resolution limit of optical microscopes. In just a few years, the first generation of transmission electron microscopes became commercially available. However, it was not until after World War II that TEMs began to be widely used for material characterization by metallurgists and materials scientists.

As shown in Figure 4.5, commercial TEM usually consists of an electron gun, a set of electromagnetic lens and apertures, a sample chamber, and a screen. Electrons beams are first generated through either field emission (monochromatic electron beam) or thermionic emission (white electrons) and then accelerated through a bias voltage to achieve high energy. The acceleration voltage determines the energy of the electron beam and thus the wavelength of the electron. Electron

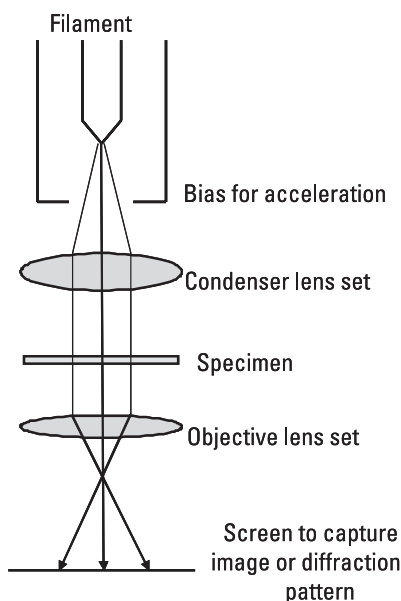


Figure 4.5 Schematic illustration of TEM including the electron gun, condenser lens set, sample/sample holder, objective lens set, and screen to capture the final image and diffraction pattern.

beams then go through the magnetic field generated by the electromagnetic lens and are thus bent and focused on the specimen. After the beam transmits through the specimen, it will go through an additional set of lens and apertures to project the information onto the screen. Here the sample preparation is critical and the specimen needs to be thin enough to be transparent for electrons. The modern technique for sample preparation includes ion milling and focused ion beam (FIB) thinning to reach the desired thickness.

TEM is widely used to observe microstructures through imaging, revealing phase/crystallographic orientation information through a diffraction pattern and discovering chemical composition by means of the energy spectrum. Like an optical microscope, a TEM can also record images of microstructures but to a much higher resolution. Typical TEM resolution can go down to the nanometer region, making it an excellent tool for nanomaterials characterization. High-resolution TEM (HRTEM) can even image lattice points that are in the range of angstroms. In fact, the current resolution of TEMs is not limited by electron wavelength but by astigmatism and aberration. A huge effort is being conducted by microscopists to resolve these issues and thus push TEM's resolution further. The transmission electron aberration-corrected microscope (TEAM) project was initiated as a collaborative effort between different national labs to redesign the electron microscope around this front [3–7]. Currently one TEAM 0.5 microscope has already been installed at the National Center of Electron Microscopy at Lawrence Berkeley National Laboratory. This double-aberration-corrected TEM is equipped with two hexapole-type spherical aberration correctors that can correct coherent axial aberration to fourth order, fifth order for spherical aberration, and sixth order for astigmatism, making this instrument capable of producing a resolution down to 0.5\AA [4–6].

Diffraction patterns can also be recorded in a TEM. According to Bragg's law, diffraction occurs when the incident beam passes through the crystal lattice. From indexing the diffraction pattern and zone axis, one can discover the phase information and the crystallographic orientations of the material under study. As mentioned in the previous section, nanomaterials will broaden X-ray diffraction peaks; the diffraction patterns from nanomaterials are also more diffused than those recorded from bulk materials. In some cases we can even record satellite patterns.

One famous example is to identify the chiral angle of a carbon nanotube (CNT). As mentioned in Chapter 2, depending on how the graphene sheet is rolled up, the carbon nanotube can have very different symmetries and thus diffraction patterns. Both the zigzag and armchair configurations have a sixfold electron diffraction pattern, but the satellite split of each chiral angle is different due to the symmetry. For example, in Figure 4.6 [8], the diffraction pattern from a SWCNT shows a strong equatorial oscillation perpendicular to the tube direction and some very weak diffraction lines from the graphite sheet elongated

TEM for CNT

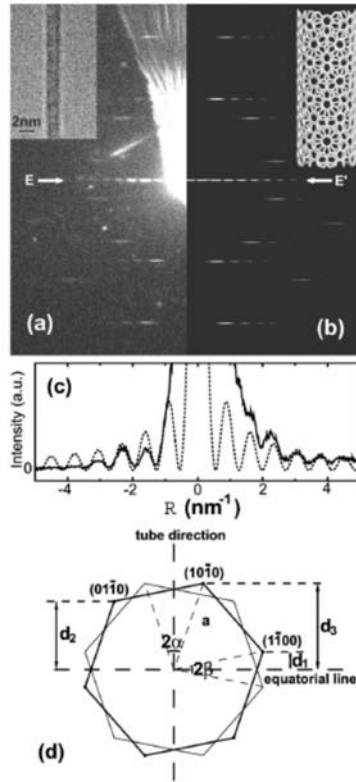


Figure 4.6 (a) A diffraction pattern from an individual SWCNT of 1.4 nm in diameter. The inset is a TEM image. (b) A simulated diffraction pattern of a SWCNT tube with a chiral vector of (14,6). The inset is the corresponding structure model. (c) Profiles of equatorial oscillation. (d) A schematic diagram of electron diffraction from an individual SWCNT. The two hexagons represent the first-order graphite-like diffraction spots from the top and bottom of the tube. (Reprinted with permission from [8]. Copyright 2003 American Institute of Physics.)

in the direction normal to the tube direction. From this diffraction pattern alone, we can determine the diameter of the CNT from the equatorial oscillation and chiral angles by measuring the distance from the diffracted lines to the equatorial line. The details can be found in [8]. This enables us to determine the CNT configuration and thus its electronic and mechanical properties.

Another interesting use of TEM is to determine the energy spectrum of specimen. As we know, some electrons are inelastically scattered by the specimen. The inelastic interaction between incident electrons and atoms inside the specimen can reveal much information about chemical composition, the electronic structure of atoms, bonding, and even dielectric properties of materials. The

spectrum, called the electron energy loss spectrum (EELS), usually consists of three different parts: zero-loss peak and low-loss and high-loss regions. The zero-loss peak consists of electrons that do not lose any energy at all. Those electrons either interact with the specimen elastically or do not interact with the specimen at all and thus will not contain much useful information. Therefore, this peak can always be ignored. The low-loss region usually represents an electron energy loss up to 50 eV. In this energy region, the interaction usually occurs with the weakly bonded outer shell electrons or inter-/intrapband transition. The high-loss region contains an energy loss greater than 50 eV and thus comes from the interaction between incident electrons with inner shell electrons. The energy absorption peak in the high-loss region is characteristic of the atom type and thus can be used for qualitative and quantitative surface chemical analyses. In addition, it is also complementary for EDAX (X-ray-based spectrum) and can be used to detect the light elements. Readers can find more details about imaging, diffraction, and the energy spectrum of TEM in many introductory TEM textbooks such as [2].

With the new developments in technologies, in situ TEM is available to researchers. Instead of standard imaging, in situ TEM records a process, such as phase transformation, deformation, or film growth, in a real time. It requires a special sample holder and chamber. For example, one can heat the sample and then cool it down to directly observe phase transformation or one can run an in situ mechanical test, such as a uniaxial tensile test or nano indentation test, and watch the deformation unfold under TEM. One can even run a creep test with in situ TEM. In situ mechanical TEM for nanostructural materials will be discussed in more detail in Chapter 10.

4.2.3 Scanning Electron Microscope (SEM)

Unlike transmission electron microscopes, a scanning electron microscope scans electron beam across the sample surface and collects scattered electrons for imaging. Because the image is formed using backscattered signals instead of forward-transmitted signals, the electron beam energy does not need to be high (<40 keV) and the sample also does not require electronic transparency; it has only a conductivity requirement to prevent charging.

As discussed earlier, at least four different kinds of signals are scattered backwards that we can potentially use: secondary electrons, backscattered electrons, and characteristic X-ray and Auger electrons. The most popular SEM imaging method is secondary electron imaging or SEI. SEI can produce good resolution imaging down to 1 to 5 nm with a large depth of field—this feature helps reveal much information about surface topography. On the other hand, backscattered electrons are often scattered by samples elastically and thus is strongly related to atomic number (Z) near the surface as well as crystallographic

orientation of the surface. Therefore, backscattered electrons are often used to identify chemical composition together with characteristic X-rays. In fact, backscattered electrons are also used to image the distribution of elements with significant difference in weight for the same reason. For example, it is almost impossible for SEI to image colloidal gold with a radius of 1 to 5 nm decorated in biospecimen, whereas backscattered electron imaging can easily distinguish these elements due to different atomic numbers. Today, backscattered electrons are also used to detect surface orientation to construct a texture map for the specimen under study. Characteristic X-rays have also been widely used in SEM to identify materials chemical composition. This is called energy-dispersive X-ray spectroscopy or EDX. As discussed previously, the wavelength of the characteristic X-ray is directly related to the energy delta between the outer and inner shells of atoms and thus is a characteristic feature of elements.

4.2.4 Scanning Probe Microscope (SPM)

Scanning probe microscopy was developed as an important technique to characterize the surface properties of materials, especially nanomaterials. The principle of SPM is very simple: A probe with very fine tip (atomic level) is used to scan a surface of interest at a very close distance, and the interaction (can be tunneling current or atomic force) between the probe and the surface is recorded. These signals are then reconstructed line by line at the atomic scale and thus reveal the topography and chemical and physical properties of the surface being studied. Depending on the signals collected, there are two types of different SPMs: One is a scanning tunneling microscope (STM), in which tunneling current between the probe tip and the surface is registered to reconstruct the surface information; the other is an atomic force microscope (AFM), in which the atomic force is recorded for a similar purpose.

The fundamental principle of STM is quantum tunneling. As discussed in Chapter 1, electrons can tunnel through a finite energy barrier if the physical distance of the barrier is small enough. As shown in Figure 4.7(a), an atomic-sharp probe tip scans cross the sample surface at a very close distance (<1 nm) with a bias applied between the probe and the sample. Under this bias, the electrons can tunnel through the small gap between the probe and the sample, and a small tunneling current can be registered through a very sensitive ampere meter. Because the current is exponentially proportional to the distance between the probe and the sample, this mechanism is very sensitive to surface topography down to the atomic level. Therefore, from the tunneling current, we can accurately measure the spatial variation and thus construct the surface image.

There are two different operating mechanisms for STM. One can either fix the probe height and measure current and thus obtain a “current” image, or keep a constant current by adjusting the distance between the probe and the

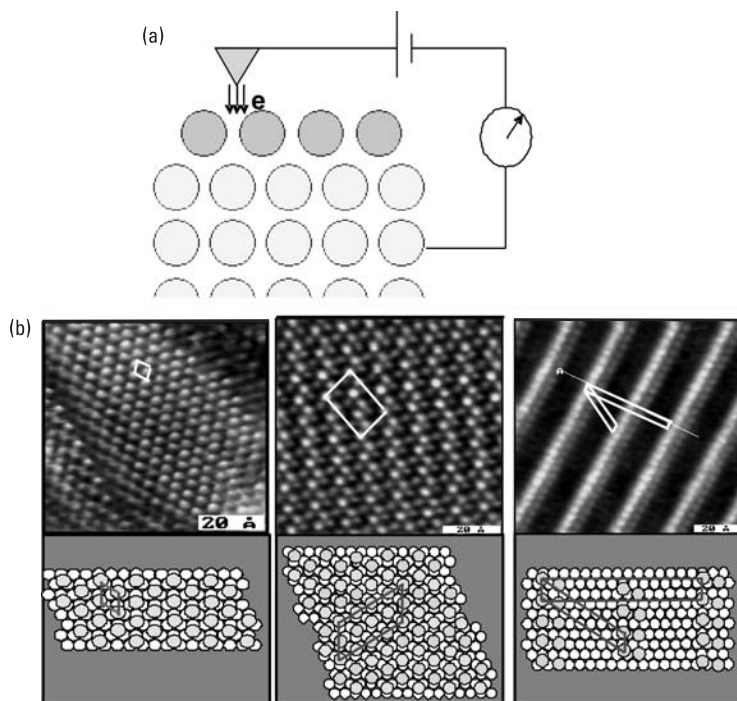


Figure 4.7 (a) Illustration of scanning tunneling microscope (STM) and (b) images obtained from STM to demonstrate atomic-level resolutions. (From unpublished work of Prof. Gangyu Liu at University of California, Davis. Used with permission.)

surface through a feedback loop controlled by ampere meters and piezoelectric ceramics. The voltage applied on piezoceramics to adjust the probe height is recorded and then used to construct the surface image. This is called a *constant current image*.

Figure 4.7(b) clearly demonstrates that the image resolution can be down to the atomic level, making STM very useful for nanomaterial characterization. In an experiment [9], decanethiol ($\text{C}_{10}\text{H}_{21}\text{SH}$, Aldrich, 98% purity) was used for a self-aligned monolayer (or SAM) on the surface of a Au(111) thin film by soaking freshly evaporated gold thin films in a 1-mM thiol solution in ethanol for at least 24 h. After that, a variable-temperature STM microscope was employed for imaging in a high-impedance, constant-current mode.

In the left panel of Figure 4.7(b), decanethiol molecules are closely packed, forming the well-known commensurate $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure with respect to the Au(111) substrate. Molecular-level defects are present at the domain boundaries. In the middle panel, $(4\sqrt{3} \times 2\sqrt{3})R30^\circ$ form a superstructure of the closely packed phase. In the right panel $(7.5 \times \sqrt{3})$ striped phase: From

the high-resolution STM topographs, the spacing between the stripes is measured to be 2.15 ± 0.10 nm, corresponding to $7.5a$ [where a is the lattice constant of Au(111)]. At high resolution, each stripe is imaged as a double row, as shown in the figure, which can be rationalized as a bimolecular row using the scaled model at the bottom. The nearest neighbor distance for the brightest spots along the stripes is 0.50 ± 0.02 nm, which is consistent with the primitive $\sqrt{3}$ structure. The observed nonidentical contrast for each double row of molecules can be explained by nonequivalent sulfur binding sites on Au(111) and the triple hollow and top sites. The phase in this panel, therefore, was determined to be the so-called $(7.5 \times \sqrt{3})$ striped phase.

Similar to STM, an atomic force microscope (AFM) also uses a probe with a fine tip scanning across the sample surface of interest. However, it does not require a tunneling current. Instead, it records the deflection of the probe tip through a piezo cantilever and constructs the image of topography while scanning through the sample surface.

Many kinds of atomic forces can be detected by an AFM. For example, if the AFM is operated under a “noncontact” mode, meaning that the probe tip is more than 1 nm away from the sample surface, van der Waals, electrostatic, magnetic, and other relatively long-range forces can be detected and used to produce images. However, under this mode, as discussed in Chapter 3, the sample surface will have to be atomically clean, meaning that there is no adsorption or any other surface defects, in order to produce images with good quality and high resolution.

The other operating mode for AFM is contact mode, in which the probe tip is less than 1 nm away from the sample surface. Under this situation, the ionic repulsive interaction is the dominant force that produces images. Additionally, this contact mode can also be used for nanomaterials patterning and fabrication, which will be discussed in Chapter 6.

4.3 Surface Analysis Methods

As discussed in Chapter 3, the surface and its related phenomena have become dominant in nanomaterials characterization. Therefore, surface analysis is an important aspect of nanomaterials characterization. In this section, we explain three different types of analysis methods that are particularly surface sensitive or surface exclusive: Auger electron spectroscopy, X-ray photoelectron spectroscopy, and secondary ion mass spectrometry. Unlike electron microscopes, these analytical methods usually will not be used to provide any spatial or structural image of a specimen, but rather to reveal the chemical composition or electronic properties such as bonding and band structures.

4.3.1 Auger Electron Spectroscopy (AES)

We introduced Auger electrons in an earlier section when we discussed the interaction between the incident electron beam and solid samples. So the primary setup for an Auger electron spectroscopy is that an incident electron beam illuminates the samples (Figure 4.8). The energy from the incident electrons helps the removal of electrons from the inner shell of sample atoms. Then electrons will fall from outer shell to the inner shell where the energy level is lower. The excessive energy is released and potentially transferred to other electrons to knock them free such that they escape from the sample. Those electrons, called Auger electrons, are detected, collected, and analyzed.

A typical AES spectrum plots electron beam intensity (or the number of electrons) versus the kinetic energy of the electrons. The kinetic energy of Auger electrons equals the energy emitted when electrons fall from the outer shell to the inner shell *minus* the energy needed for an Auger electron to escape from the surface. For example, if the incident beam kicks out an electron from the inner K shell, then an electron from the L shell falls into the K shell. The energy emitted from this process is $E_K - E_L$. This energy then can be used to release the other Auger electrons with E_A . So the final kinetic energy the detector recorded is $E = E_K - E_L - E_A$. Because all of these three types of energy are element specific, the Auger spectrum is very characteristic and has been widely used for surface chemical composition analysis.

4.3.2 X-Ray Photoelectron Spectroscopy (XPS)

An X-ray photoelectron spectroscopy is very similar to an AES. The setup is almost identical to that of an AES except that the energy source is a monochromatized X-ray instead of an electron beam. The photons with well-defined energy are absorbed by the specimen, causing the emission of the electrons. These electrons are called X-ray photoelectrons.

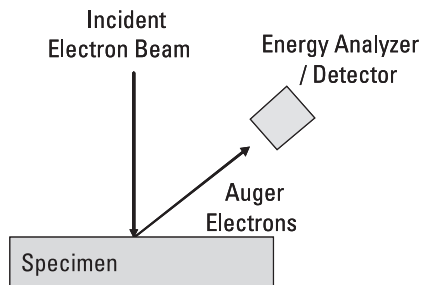


Figure 4.8 Illustration of AES. In the case of XPS, the difference is that the incident beam is X-ray.

The XPS is also similar to the AES spectrum, plotting electron intensity versus electron kinetic energy. However, unlike AES where the energy of Auger electrons is the characteristic feature of material elements, the kinetic energy of XPS electrons is actually closely related to the binding energy of materials. If the incoming X-ray photon's energy is E_i and the kinetic energy of photoelectrons is E_p , the binding energy can be calculated as $E_b = E_i - E_p$. Therefore, one can directly obtain binding energy information and subsequent bonding/electronic band and even density of state information from XPS.

4.3.3 Secondary Ion Mass Spectroscopy (SIMS)

Secondary ion mass spectroscopy is another surface chemical analysis method. However, this method is very different compared to AES and XPS. First, the incident energy source now is an ion beam (usually heavy metals such as Cs, W, or Ga) that is accelerated with an energy range of 1 to 30 keV. When the incident beam hits the sample surface, it sputters the surface, kicking out the particles that can be either charged positive or negative or neutral. Those particles then pass through an energy filter and are collected in a mass spectrometer. The typical spectrum plots ion yield or intensity versus mass number and thus determines the surface chemical composition.

SIMS can be used in a scanning mode and a spatial image of surface chemical composition can be constructed. In addition, SIMS can also be used in a sputtering mode where chemical composition depth profile can be obtained as new surface is exposed and examined by sputtering.

References

- [1] Warren, B. E., *X-Ray Diffraction*, Mineola, NY: Courier Dover Publications, 1990.
- [2] Williams, D. B., and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, New York: Springer, 1996.
- [3] Dahmen, U., "A Status Report on the TEAM Project," *Microscopy and Microanalysis*, Vol. 13 (Suppl. 2), 2007.
- [4] Kisielowski, C., et al., "Detection of Single Atoms and Buried Defects in Three Dimensions by Aberration-Corrected Electron Microscopy with 0.5 Å Information Limit," *Microscopy and Microanalysis*, Vol. 14, 2008, pp. 454–462.
- [5] Meyer, J. C., et al., "Direct Imaging of Lattice Atoms and Topological Defects in Graphene Membranes," *Nano Letters*, Vol. 8, No. 11, 2008, pp. 3582–3586.
- [6] Kisielowski, C., R. Erni, and Frertag, B., "Object-Defined Resolution Below 0.5 Å in Transmission Electron Microscopy—Recent Advances on the TEAM 0.5 Instrument," *Microscopy and Microanalysis*, Vol. 14 (Suppl. 2), 2008, pp. 78–79.

- [7] Dahmen, U., et al., “An Update on the TEAM Project—First Results from the TEAM 0.5 Microscope, and Its Future Development,” *Instrumentation and Methods*, T. Duden et al., Eds., 2008, Vol. 1, 2008, pp. 3–4.
- [8] Gao, M., et al., “Structure Determination of Individual Single-Wall Carbon Nanotubes by Nanoarea Electron Diffraction,” *Applied Physics Letters*, Vol. 82, No. 16, 2003, pp. 2703–2705.
- [9] Liu, G., et al., unpublished research work, University of California, Davis, 2009.

Part II

Nanomaterials Fabrication

Part II of this book covers fabrication of nanostructures. Methods used to generate nanoscale structures and nanoscale-structured materials can be divided roughly into two groups: “top-down” and “bottom-up” approaches. The top-down approach starts with large objects at least in one or two dimensions and reduces their lateral dimensions in order to achieve fine feature and nanoscale materials. This approach starts with thin-film deposition and then applies serial and parallel techniques for patterning features—typically in two dimensions—over the length scale, approximately 4 orders of magnitude larger than an individual structure. A bottom-up approach produces nanoscale structures, such as nanoparticles, as small building blocks and then assembles them into larger nanostructures for desirable applications. In this part, we introduce two important steps of the top-down approach—thin-film deposition and nanoscale patterning in Chapters 5 and 6, respectively—and then a bottom-up approach in Chapter 7.

5

Thin-Film Deposition: Top-Down Approach

To obtain nanoscale structures, a top-down approach is used to first fabricate a large object that is 2 to 3 orders larger in one or two dimensions than the nanoscale desired and then nanopatterning techniques are utilized to achieve smaller features. This method actually was first developed by and has been widely used in the microelectronics industry. Now with more advanced thin-film deposition and nanopatterning methods, this approach has been pushed further into the regime of nanofabrication. We focus on thin-film growth mechanisms and deposition methods in this chapter and introduce advanced nanopatterning techniques in the next chapter.

5.1 Thin-Film Deposition Mechanisms

Thin-film growth or deposition involves both films to be deposited and substrates. Depending on the relationship between the film and the substrate, thin-film growth can be divided into homogeneous and heterogeneous film growth. Homogeneous film growth is the growth or deposition of a film onto a single crystalline substrate of the same materials and crystal orientations, whereas heterogeneous thin films grow or deposit on a substrate that consists of different materials. We discuss these respective deposition mechanisms in the following sections.

5.1.1 Homogeneous Film Growth Mechanisms

Because a homogeneous thin film is grown or deposited on a substrate with the same orientation, there is no lattice mismatch. Hence, the total Gibbs energy

is dominated by the surface energy. Even a well-polished, single-crystal substrate will still have the surface structures, and thus the growth of a thin film on top of the substrate is favorable thermodynamically if the surface energy (or equivalently the surface area in this case) can be minimized. However, as mentioned earlier, a thermodynamic equilibrium state may not be achievable depending on growth kinetics. Here we introduce two mechanisms for homogeneous thin-film growth based on the balance between thermodynamics and kinetics.

5.1.1.1 Step Propagation

As shown in Figure 5.1, although smooth microscopically, any single-crystal surface consists of a terrace, steps, and kinks. If we use a little cube to represent an atom randomly added to the surface during homogeneous film growth, we can determine which position this atom will go thermodynamically. If it is added on the terrace, it stands out like an island with four cubic side walls as extra surface. However, the extra surface is only two side walls if the atom grows from the steps and 0 (meaning no extra surface energy is added) from the kinks.

Therefore, if we want to minimize the overall surface area, we would like the atom to land on and grow from kinks. This is called the *step propagation mechanism*. However, we now know that growth from kinks is thermodynamically preferred but the probability of a random atom landing on the kinks is not very high—atoms tend to land on large areas such as the terrace, as evidence by simple probability calculations. Therefore, the atoms landing on other places

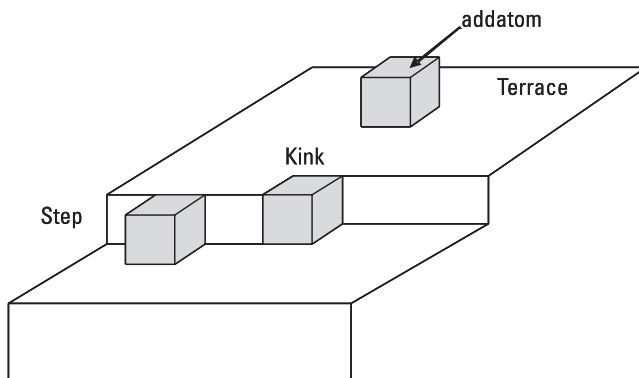


Figure 5.1 Illustration of homogeneous thin-film growth mechanisms. When an addatom lands on the terrace, it will create extra surface area that is equivalent to four cubic side walls. However, if the addatom can move to steps and kinks due to large surface diffusion at high temperature, the extra surface area created will be minimized. However, at lower temperature, surface diffusion is not high so the addatom will form a cluster on the terrace and grow film in an island mechanism.

will have to go through some lateral movements to get to the kinks through surface diffusion. As a consequence, the step propagation mechanism of homogeneous film growth is dominated at high temperatures when surface diffusion is fast enough that randomly landed atoms can diffuse to those thermodynamically favorable places such as kinks and steps.

5.1.1.2 Island Growth

On the other hand, if surface diffusion is slow, kinetically there is not enough time for atoms to move to those thermodynamically favorable places such as kinks and steps. Instead, if the deposition rate is high, the atoms that landed on the surface may encounter each other easily on the terrace and, thus, form a cluster of atoms to minimize the surface energy since the cluster has less exposed surface area than the sum of the surface areas of all of the individual atoms. However, once a cluster has formed, it is even less mobile than the individual atoms and, therefore, more stable and will attract more atoms to form an island—thus, the island growth mechanism occurs at low temperatures where surface diffusion is low, but the deposition rate is high.

5.1.2 Heterogeneous Film Growth Mechanisms

Unlike homogeneous film growth, heterogeneous films are grown or deposited on the substrate of different materials and structures. Thus, the total energy includes both the surface energy and the elastic energy due to lattice mismatch. Therefore, the overall growth mechanisms depend on the competition and balance of these two energy terms. We can classify them into the three categories discussed next and shown in Figure 5.2.

5.1.2.1 Frank–van der Merwe Model

If the lattice parameters of thin films match the substrate, even though they may be different materials, the elastic energy is negligible. So the total energy depends on the surface energy. However, now that the thin film and the substrate are different materials, the surface energy needs to include more terms. To form a continuous film, we should have film spreading on the surface to form a continuous layers—this is similar to the wetting criteria we discussed in Chapter 3.

As shown in Figure 5.2(a), if the surface energy of the substrate is equal to or higher than the sum of the surface energy of film and the interface energy between film and substrate, the film can completely wet the substrate and grow a layer-by-layer thin film. This is called the Frank–van der Merwe model.

5.1.2.2 Volmer–Weber Model

On the other hand, if the film cannot completely wet the substrate, or the substrate surface energy is smaller than the sum of the film surface energy plus

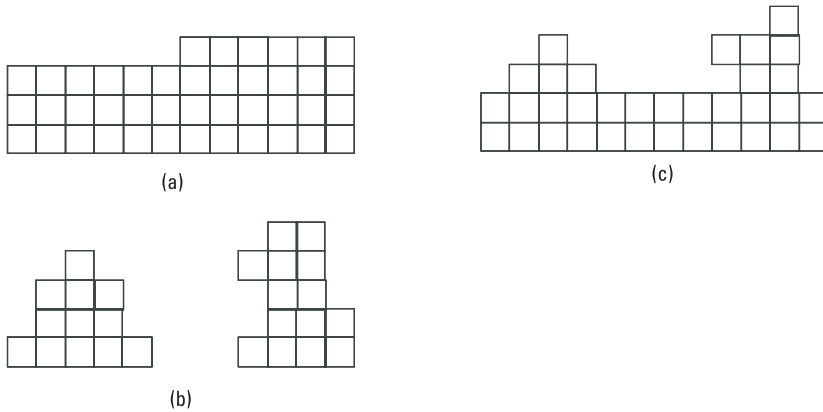


Figure 5.2 Illustration of heterogeneous thin-film growth mechanisms: (a) layer-by-layer growth through Frank–van der Merwe model, (b) island growth through Volmer–Weber model, and (c) a mix mode through Stranski–Krastanov Model.

the interface energy between film and substrate, the film will not tend to spread itself on the substrate. Therefore, it will not form layer-by-layer continuous growth. Instead, it will grow into a 2D island, which is called the Volmer–Weber model [shown in Figure 5.2(b)].

Although this is also 2D island growth, the Volmer–Weber model is different from the one we discussed in homogeneous film growth where the 2D island is formed purely because of kinetic reasons. The Volmer–Weber model of 2D island growth in heterogeneous film is a result of balancing surface/interface energy, so it is a result of thermodynamics.

5.1.2.3 Stranski–Krastanov Model

Both the Frank–van der Merwe and Volmer–Weber models assume a perfect lattice match. However, in reality, we know that it is very difficult to match the lattice of films with substrates exactly. Therefore, most of the heterogeneous film growth will have to take the elastic strain energy into consideration. This is done in the Stranski–Krastanov model.

As shown in Figure 5.2(c), this model is based on the competition between the surface energy and the elastic strain energy. For surface energy, it assumes the film will completely wet the surface as it does in the Frank–van der Merwe model, so the surface energy is proportional to the film surface area. However, the elastic energy accumulates as each layer grows, so it is proportional to the total volume. Therefore, at the initial stage, the film is very thin; the surface energy is the dominant factor, while the elastic energy is almost negligible.

At this time, the films will follow layer-by-layer continuous growth just as in the Frank–van der Merwe model. However, as the film thickness increases,

the elastic strain energy due to lattice mismatch starts to increase and grow at a much faster rate and gradually become the dominant term. Once the thickness reaches a critical value (normally called critical film thickness), the strain energy becomes larger than the surface energy and the film can no longer continue to grow layer by layer. Instead, it will switch to island growth in order to minimize the elastic strain energy.

Therefore, critical film thickness is the boundary below which the surface energy is dominant and thus layer-by-layer continuous growth is preferred, and above which, the elastic energy is dominant and the film tends to form 2D island growth. In other words, if we want to form continuous film, we will have to keep the film thickness below the critical film thickness as shown in Figure 5.3.

5.2 Thin-Film Deposition Methods

There are many ways to deposit thin films on substrates. For example, plating deposition of a metallic film uses a solution to transport metallic ions from the anode (target) to the cathode (substrate), which can be connected through an external circuit with (electroplating) or without current (electroless plating). However, the most popular method that has been widely used in nanofabrication is vapor deposition. In this section we focus on this thin-film deposition method, which can be further divided into two categories: physical vapor deposition and chemical vapor deposition. Readers can find more details in [1].

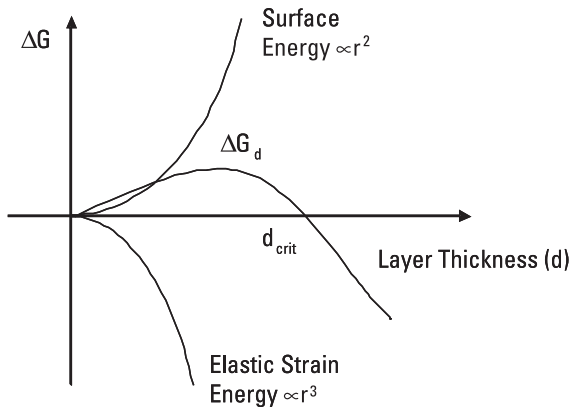


Figure 5.3 Illustration of competition between surface energy and elastic energy during heterogeneous film growth. As a consequence, there exists a critical thickness below which film can grow layer by layer, and above which film can no longer continuously grow. Instead, it will break down to several islands in order to relax the elastic energy.

5.2.1 Physical Vapor Deposition (PVD)

Physical vapor deposition is one type of vacuum deposition method that utilizes the condensation of vapor on the substrate surface. The process does not involve a chemical reaction. In general, this method involves a substrate, which can be heated or biased for better process control (such as uniformity); a target containing the material to be deposited onto the substrate; and a source of energy to convert the target material into vapor. The process usually happens in high vacuum.

Based on the different sources of energy, PVD can be subdivided into different methods, such as evaporation, in which thermal energy is used to heat and evaporate the target; sputtering, in which plasma is used to sputter the target and direct ions to redeposit onto the substrate, electron beam, or ion beam; or laser deposition, in which the energy source to sputter target is either an electron beam, ion beam, or pulsed laser.

5.2.1.1 Thermal Evaporation

Thermal evaporation consists of two steps: Material is evaporated from a hot source and then condenses on a relatively cold surface. The most commonly observed thermal evaporation is the evaporation and condensation of water on the lid of the cooking pot.

In film deposition, thermal evaporation usually takes place in a high-vacuum system. Because the energy source is purely thermal energy, the equipment setup is relatively simple. The deposition species are atoms and ions—usually these atoms and ions do not collide with each other on their journey from the source to the substrate due to high vacuum. On the other hand, the substrate is often not biased, thus the kinetic energy for the deposition species usually is low (0.1 to 0.5 eV). The advantages of this method are the high deposition rate and large substrate size. The downside is that the materials to be deposited are limited to those with low sublimation or evaporation temperatures.

Another method that is similar to thermal evaporation but much more sophisticated is molecular beam epitaxy (MBE). Just as the name suggests, MBE is a type of epitaxy growth, meaning that the film and the substrate share the same or closely related orientations. This requires a slow deposition rate and thus an ultrahigh-vacuum environment. In a typical solid source MBE system, the target, such as Si, Ge, Ga, or As, is heated in a separate cell until sublimation occurs. The vapor is then introduced into the main deposition chamber and condenses on the substrate. Due to the requirement for an ultrahigh vacuum and ultraclean substrate surface, the equipment system is usually very complicated and the deposition rate is very slow. However, the films deposited by MBE are usually high-quality epitaxy films on single-crystal substrates.

There are other ways to evaporate materials, such as pulse laser deposition or electron-beam deposition. In these cases, the input energy itself is not thermal

energy. But a pulse laser or the narrow beam of a high-energy electron beam will lead to localized heat at the incident point and thus evaporate source materials. Ultimately, the evaporation mechanism is still thermal energy, although the input energy is not. These methods have been widely used for some refractory materials where normal joule-heating evaporation will not work due to high melting points. In addition, because it uses a laser and electron beam, the energy source and the equipment can be very expensive and the substrate will also have to be limited to small size due to the vacuum system.

5.2.1.2 Sputtering

Whereas the mechanism for evaporation is thermal energy, sputtering is based on momentum transfer. During sputtering, the high-energy ions bombard the target surface and transfer an amount of kinetic energy that far exceeds the chemical bonding strength of the target materials. Therefore, the source material is “decomposed” or “sputtered” from the target and deposited on the substrate. Normally the materials to be deposited are metals (Al, Cu, Ti, W, and the like) or metal-like compounds (TiN, TaN).

The primary source of ion bombardment is either plasma or high-energy ion gas. Plasma, often called the fourth state of materials, is partially ionized gas consisting of positive ions and negative electrons. The positive ions are accelerated and hit the surface of the target, causing a collision cascade inside the target materials and eventually leading atoms sputtered out of the target through momentum transfer.

Due to different bonding energies, different elements have different sputtering yields. However, when the target is an alloy, if one element is preferentially sputtered off the surface of the target, the surface concentration of the other element will become higher and, thus, ion bombardment will have more chances to sputter the other element. Therefore, sputtering will eventually reach a quasiequilibrium and deposit films with the same stoichiometry as the target.

There are several ways to generate plasma for sputtering; for example, dc sputtering, magnetron sputtering, and RF sputtering. As shown in Figure 5.4,

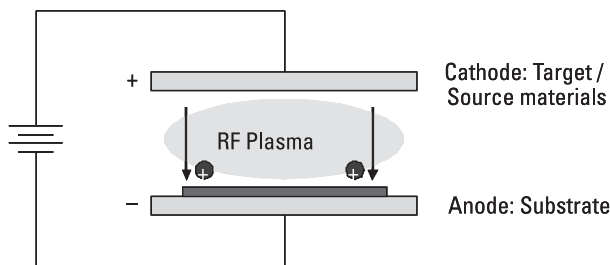


Figure 5.4 Schematics of dc sputtering PVD process.

dc sputtering is accomplished by applying a dc voltage across two parallel plates with the substrate as anode (–) and the target as cathode (+). When the voltage applied is high enough, plasma will be generated. Positive plasma ions will sputter the surface and knock off source material out of the target as positive ions and neutral atoms and deposit on the substrate at the anode.

DC sputtering is a very simple and easy way to accomplish sputtering and works very well as long as the substrate and target are conductive. For an insulating target, a high RF plasma is usually applied across the anode and cathode through a capacitive coupling. At the same time, the substrate/anode is grounded together with the chamber wall to form a nonsymmetric plasma and ensure deposition on the substrate. Nevertheless, the deposition rate for RF plasma sputtering is much slower than that of dc sputtering. To improve the sputter deposition rate, one can increase the ionization of plasma by applying additional magnetic field. This is so-called “magnetron sputtering,” which has been widely used in semiconductor manufacturing. In a magnetron system, electrons are forced to rotate in a helical path near the cathode and thus have a much higher chance for ionization of plasma.

5.2.2 Chemical Vapor Deposition (CVD)

Unlike the PVD method, chemical vapor deposition involves a chemical reaction at the substrate surface in order to form the thin film. As shown in Figure 5.5, a typical CVD process involves at least five steps: (1) mass transportation of reactants delivered from the gas/liquid source to the substrate surface; (2) the adsorption of reactants on the substrate surface; (3) the chemical reaction on the substrate surface to form the desired film; (4) desorption of by-products of the chemical reaction from the substrate surface; and (5) pumping away of the by-products and unreacted reactant. The deposition rate of a CVD process is basically controlled by the slowest step in this serial process.

Among these five steps, adsorption and desorption usually occur relatively fast. Therefore, the deposition rate is mainly determined by the competition between mass transportation (of either reactants in step 1 or by-products in step 5) and the surface reaction rate in step 3. Mass transportation is usually a function of gas flow and its partial pressure while surface reaction rate is exponentially proportional to temperature.

At high temperatures, the surface chemical reaction is dominant. Precursor will react and form the thin film as soon as it reaches the substrate surface. Therefore, this process is controlled by mass transportation. The mass transportation can be increased by increasing the supply of reactant such as gas or liquid precursor, thus increasing the deposition rate. The deposition usually starts at multiple nucleation sites and grows and coalesces together. Because the deposition rate is strongly influenced by mass transportation, the film tends not to

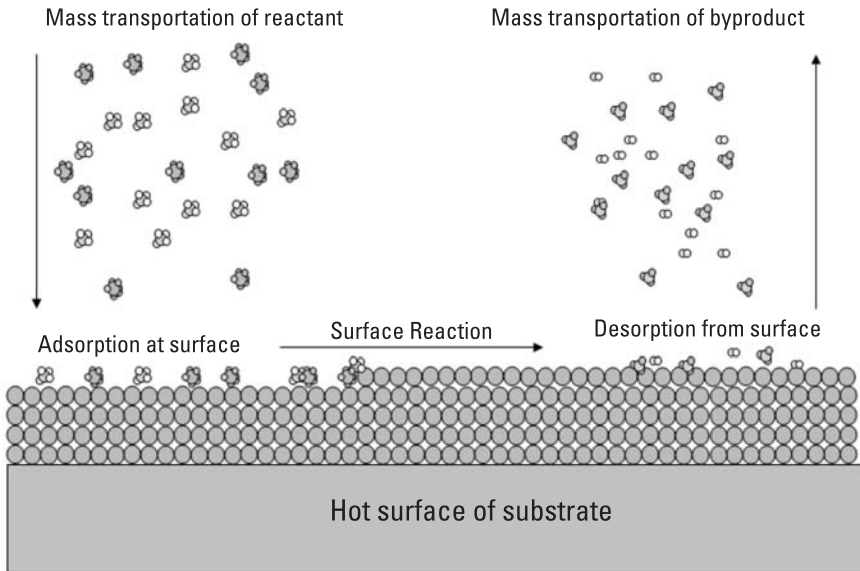


Figure 5.5 Illustration of a typical CVD process involving five steps: (1) mass transportation of reactants delivered from the gas/liquid source to the substrate surface; (2) the adsorption of reactants on the substrate surface; (3) the chemical reaction on the substrate surface to form the desired film; (4) desorption of by-products of the chemical reaction from the substrate surface; and (5) pumping away of the by-products and unreacted reactant.

be conformal if the surface has pronounced topography and usually it is also difficult to control the deposition rate.

On the other hand, for a low-temperature regime, the surface reaction rate is slow and becomes the dominant step. In this case, regardless of any reactant flow change, the deposition rate will remain the same. Due to the slow reaction rate, the surface will have sufficient reactant supply, and reactions usually happen at all surface areas simultaneously, leading to a very uniform and conformal deposition. However, if the temperature is too low, the reaction may not occur at all for thermodynamic reasons. So, choosing the right temperature is critical for the CVD process.

5.2.2.1 Normal CVD Reactant

A normal CVD reactant can be either gas or liquid. There can also be more than one reactant. If there is only one reactant in the CVD process, it usually has a thermal decomposition reaction on the substrate surface. When more reactants are involved, they usually include a precursor (SiH_4 , WF_6) and a reducing agent (NH_3 , H_2). The normal CVD reactant temperature can vary

but normally should be relatively high due to the fact that the precursors need to be disassociated.

5.2.2.2 Plasma-Enhanced CVD (PECVD)

In many modern thin-film deposition techniques, dc or RF plasma is used in the CVD reactor, just like in the PVD process. In this case, the plasma will help break down the precursors before they reach the substrate surface, thus significantly decreasing the surface temperature required for chemical reactions. However, high vacuum (in the range of torrs to millitorrs) is required in order to ignite the plasma.

5.2.2.3 Atomic Layer Deposition (ALD)

Atomic layer deposition utilizes the fact that at low temperature, the deposition rate is determined by the surface reaction and independent of mass transportation. Each cycle of ALD consists of the following steps (as shown in Figure 5.6):

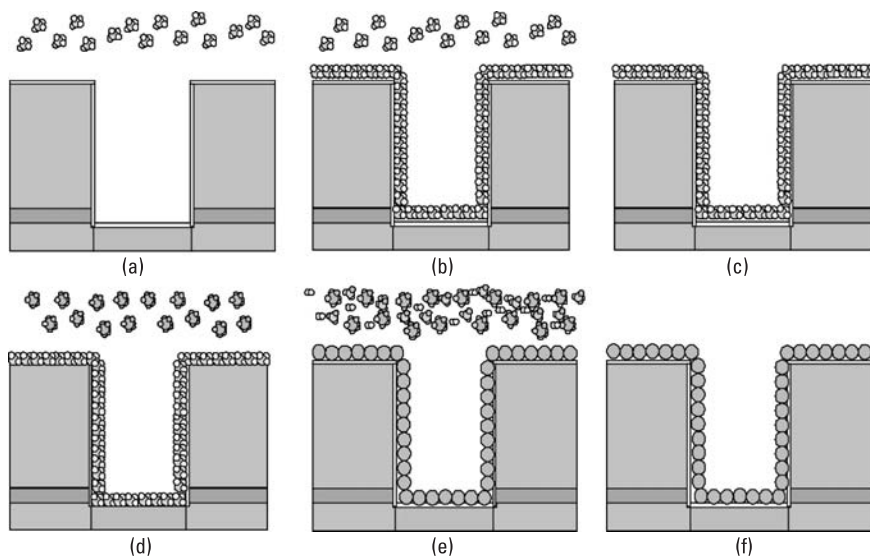


Figure 5.6 Schematics of one cycle of ALD: (a) Introduce reactant 1 into chamber; (b) reactant 1 completely adsorbed on substrate surface; (c) pump away the residual reactant 1; (d) introduce reactant 2 into chamber; (e) reactant 2 will react with monolayer of reactant 1 and form an atomic layer of the desired film on the substrate and some by-products; (f) pump away the residual reactant 2 and any by-product. Standard ALD process usually consists of dozens of cycles repeating from steps a to f.

- a. Introduce reactant 1 into chamber.
- b. Reactant 1 completely adsorbed on substrate surface.
- c. Pump away the residual reactant 1.
- d. Introduce reactant 2 into chamber.
- e. Reactant 2 will react with monolayer of reactant 1 and form an atomic layer of the desired film on the substrate and some by-products. This surface reaction is self-limited due to availability of reactant 1.
- f. Pump away the residual reactant 2 and any by-product.
- g. Repeat steps a to e to start next cycle until desired film thickness is achieved.

Due to the slow deposition rate, each cycle of ALD actually deposits only a fraction of the atomic layer; in other words, several cycles will produce a one-atomic-layer-thick film, making it perfect technology for thickness control in nanofabrication. Because it is surface reaction limited, it usually has very good conformality, making it very appealing for nanoparticle fabrications. However, the reaction has to be at a low temperature, so choosing the appropriate precursors is very critical for this method.

Table 5.1 summarizes the thin-film deposition techniques introduced in this section.

Table 5.1
Summary of the PVD and CVD Thin-Film Deposition Techniques

	Physical Vapor Deposition			Chemical Vapor Deposition
	Evaporation/ MBE	Sputtering	Pulsed Laser Deposition	CVD/PECVD/ ALD
Mechanism	Thermal energy	Momentum transfer	Laser-induced thermal energy	Chemical reaction
Deposition rate	High, up to 750,000 Å/min	Low, except for pure metals	Moderate	Moderate, up to 2,500 Å/min
Deposition species	Atoms and ions	Atoms and ions	Atoms, ions, and clusters	Precursor molecules dissociate into atoms
Energy of species	Low; 0.1–0.5 eV	Can be high; 1–100 eV	Low to high	Low; can be high if plasma enhanced
Substrate size	Large	Large	Limited	Large

Reference

- [1] Smith, D. L., *Thin Film Deposition: Principles and Practices*, Columbus, OH: McGraw-Hill Professional, 1995.

6

Nanolithography: Top-Down Approach

In Chapter 5, we introduced thin-film deposition as the first part of a top-down approach for nanofabrication. In this chapter, we discuss how to use nanolithography to pattern the thin film we deposited to a small scale. We focus mainly on the most important and widely used techniques. After a brief introduction to the most commonly used nanolithography techniques in Section 6.1, we focus on several promising approaches in Sections 6.2 through 6.4. Following this, in Section 6.5 we discuss a few top-down techniques that are accomplished with the assistance of a bottom-up approach.

6.1 Introduction

In the last decade, nanolithography has been used to advance areas of modern science and technology such as the production of integrated circuits, information storage devices, video screens, microelectromechanical systems (MEMS), miniaturized sensors, microfluidic devices, biochips, photonic bandgap structures, and diffractive optical elements [1–4]. Generally, nanolithography can be divided into two categories based on the patterning strategy used: parallel replication and serial writing. Parallel replication methods such as photolithography [5], contact printing [6], and nanoimprint lithography [7] are useful for high-throughput, large-area patterning. However, most of these methods can only duplicate patterns, which are predefined by serial writing approaches and thus cannot be used to arbitrarily generate different patterns (that is, one photomask leads to one set of feature sizes for a given wavelength). In contrast, serial writing methods, such as scanning probe lithography (SPL) [8], can create patterns

with high resolution and precise registration, but they are limited in throughput [9].

6.1.1 Parallel Replication

Parallel replication methods duplicate predefined patterns, which are the mask reticles in photolithography or the molds in contact printing and nanoimprint lithography. Photolithography is the oldest top-down method. It uses light shining through the masked area on a photoresist coated substrate (a Si wafer). The substrate is covered by a photoresist, which is the key to the photolithography process. The photoresist is an organic-based compound. There are two types of photoresists: positive resists and negative resists. Positive resists are stripped away and negative resists remain after exposure to light. In the case of a positive resist, the substrate is exposed to the light and the exposure area without a mask is removed. The substrate is etched either in plasma or in solvent to remove certain areas and make patterns on the substrate. Figure 6.1 shows a general illustration of photolithography.

The limitation of the photolithography technique is the dimension of patterns. The size of pattern that can be produced is up to the wavelength of the light, which limited to about 37 nm. The process can be done only on a hard flat surface and in a provided facility. Due to this limitation, several studies have been conducted to find an alternative way to transfer micropatterns on different substrates.

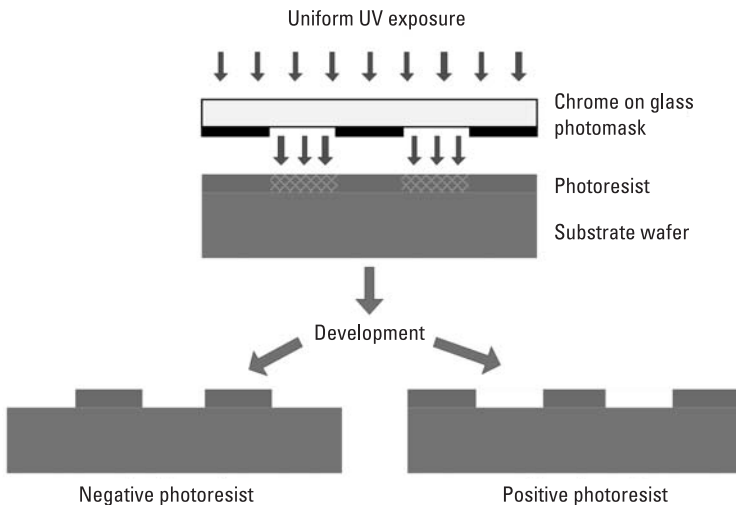


Figure 6.1 Schematic of the photolithography fabrication process.

Microcontact printing is a well-known method for transferring molecules from a master to a substrate to create well-defined patterns by physical contact. This method was first described in 1993 by Kumar and Whitesides [6], who reported using a patterning elastomeric stamp with hexadecanethiol as an ink brought into contact with a Au-coated surface. The hexadecanethiol molecules form a monolayer in the area of contact between the elastomeric stamp and the Au surface area. The patterned hexadecanethiol molecules protect the Au surface from etching in an aqueous, basic solution of cyanide ion and dissolved dioxygen (1 M KOH, 0.1 M KCN, and saturated O₂). After etching, hexadecanethiol molecules generate well-defined patterns with dimensions ranging from microns to centimeters. Figure 6.2 shows the schematics for the fabrication of Au patterning using an elastomeric stamp and hexadecanethiol molecules as an ink.

The combination of aqueous, alkaline cyanide ion and dioxygen rapidly removes unprotected Au(0) areas, and forms a stable cyanide complex of Au in solution [10]:

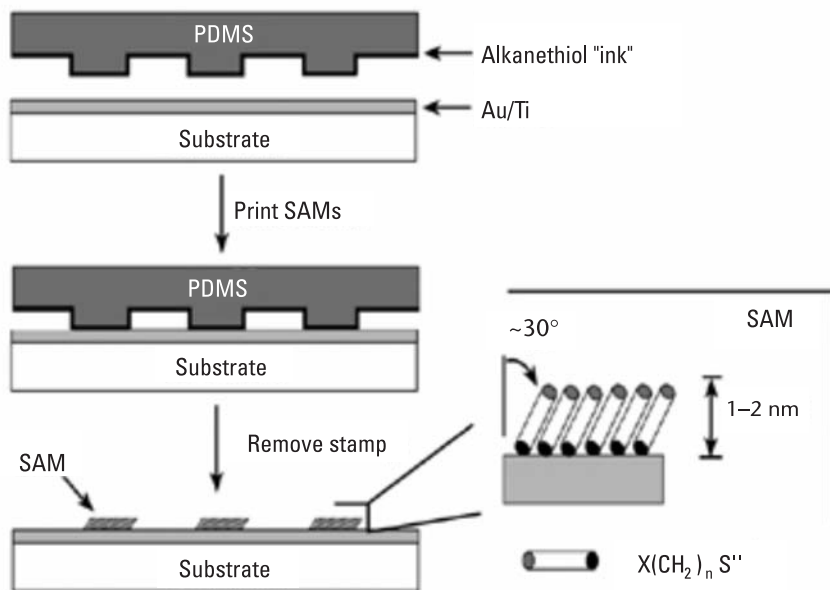
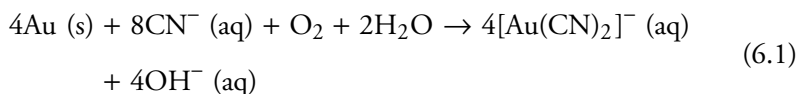


Figure 6.2 Schematic illustrating the transfer of alkanethiol molecules from an elastomeric stamp to a substrate. (Reprinted with permission from [6]. Copyright 1993 American Institute of Physics.)

The sulfur part of hexadecanethiol molecules is chemisorbed onto an Au surface and forms a strong covalent bond. Alkyl chains of hexadecanethiol molecules create van der Waals forces and block cyanide anions coming into the Au surface area. Tremendous efforts are being aimed at investigating areas of application, the transfer mechanism, and extensions of substrates and inks.

Nanoimprint lithography is another parallel patterning method. It is discussed in detail in the next main section.

6.1.2 Serial Writing

Unlike parallel methods, a predefined pattern is not needed when using serial writing. Arbitrary patterns can be generated via this method. However, this method is not suitable for throughput. Several serial writing approaches are discussed in Section 6.3.

6.2 Nanoimprint Lithography (NIL)

Nanoimprint lithography is a high-throughput, high-resolution parallel patterning method in which the surface pattern of a stamp is replicated into a material by mechanical contact and 3D material displacement. After the first article about it was published in 1995 [7], NIL was quickly recognized by researchers and industry as a potentially low-cost, high-throughput lithographic method for a range of applications. In 2003 NIL was accepted by ITRS as a next generation lithography candidate and found its way to the road map for the 32-nm node and beyond, scheduled for industrial manufacturing in 2013.

6.2.1 NIL Process

Imprint lithography has two steps: imprint and pattern transfer (Figure 6.3). In the imprinting step, a mold with nanostructures on its surface is used to deform a thin resist film deposited on a substrate. In the pattern transfer step, an anisotropic etching process, such as reactive ion etching (RIE), is used to remove the residual resist in the compressed area, transferring the thickness contrast pattern created in the mold into the entire resist. The resist can be a thermal plastic (hot embossing NIL) or UV curable polymer (UV-NIL) or other deformable material. For a thermal plastic resist, the resist is heated above its glass transition temperature during imprinting, such that the resist becomes a viscous liquid and the mold is pressed into the resist. Then the resist is cooled below its glass transition temperature before the mold is removed.

NIL is a primarily a physical deformation process, which, although simple, allows NIL to avoid many problems in other lithographical methods and to

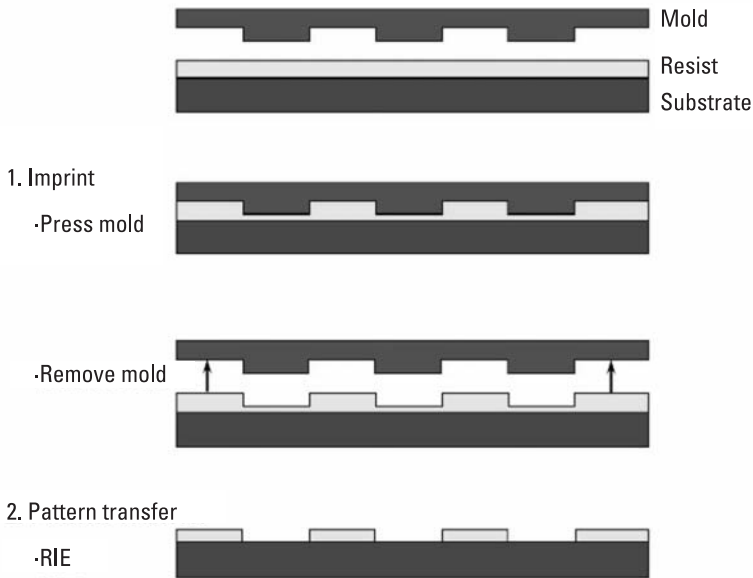


Figure 6.3 Schematic of nanoimprint lithography process: (1) imprinting using a mold to create a thickness contrast in a resist, and (2) pattern transfer using anisotropic etching to remove residue resist in the compressed areas.

achieve high resolution and high throughput and low cost at the same time. NIL has demonstrated the ability to pattern 5-nm-line-width and 6-nm-half-pitch features (see Figure 6.4) [11, 12].

6.2.2 3D Patterning via NIL

Rather than the 2D patterning used in conventional lithography, 3D patterning is available via NIL [13]. Three-dimensional features are very desirable for certain applications such as microwave circuits and MEMS. For example, the T-gate for microwave transistors has a narrow footprint for high-frequency operation but a wide top for lower resistance. Fabrication of a T-gate often requires two electron-beam lithography steps: one for the narrow footprint and one for the wide top. Each electron-beam exposure could take more than 2h to pattern a single 4-in. wafer. With NIL, the entire 4-in. wafer can be patterned in one step in less than 10 sec. Figure 6.5 shows two T-gates of 40- and 90-nm footprint fabricated by a single NIL step and liftoff of metal.

6.2.3 Air Cushion Press

Imprint pressure uniformity is crucial to the pattern uniformity and yield of nanoimprint lithography and, hence, its applications. NIL creates patterns by

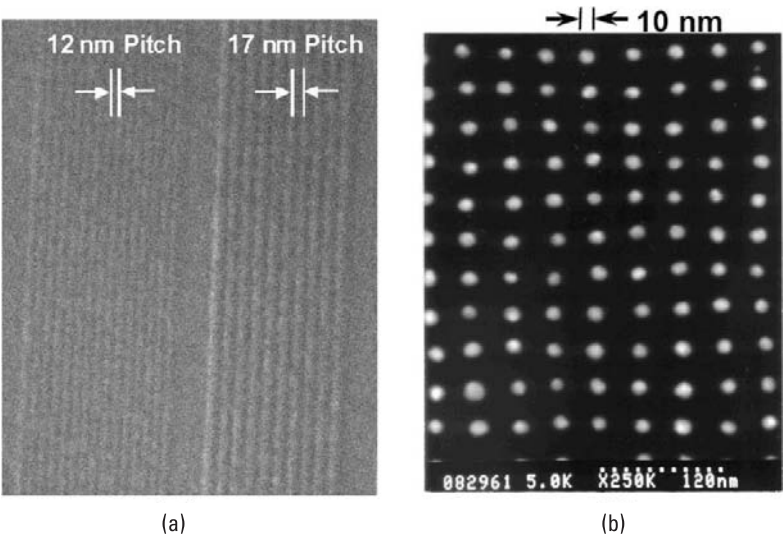


Figure 6.4 Scanning electron microscope (SEM) images of (a) imprinted resist grating with a minimum 6-nm half-pitch and (b) a 10-nm diameter and 40-nm period of metal dot array by nanoimprint and a liftoff. [(a) Reprinted with permission from [11]. Copyright 2005 Institute of Physics. (b) Reprinted with permission from [12]. Copyright 1997 American Institute of Physics.]

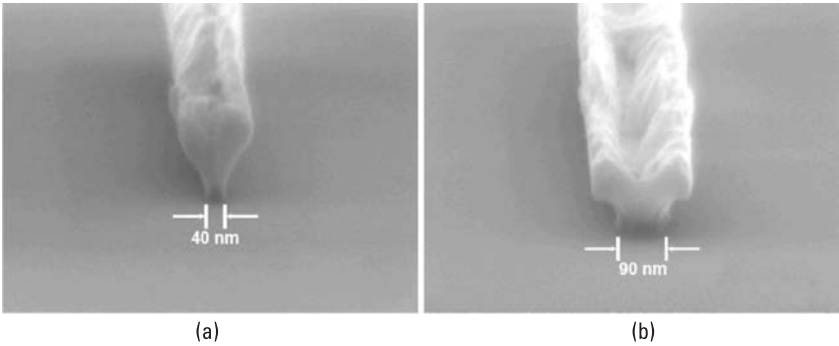


Figure 6.5 Three-dimensional patterning. SEM images of two T-gates of (a) 40- and (b) 90-nm footprint fabricated by a single NIL and a liftoff of metal. (Reprinted with permission from [13]. Copyright 2001 American Institute of Physics.)

physically deforming a material (a polymer resist or a functional material) with a mold; hence, it requires intimate contact between the mold and substrate, which, in turn, requires a uniform imprint pressure everywhere over the entire imprint field. Moreover, to reduce damage to the mold and prolong mold

lifetime, one should avoid high-pressure points over the imprint field and any relative rotation and lateral shifting between the mold and substrate during the imprint process.

To meet these requirements, a new technique, air cushion press (ACP) (Figure 6.6), has been developed [14]. ACP utilizes a gas (or fluid) to press the mold and substrate against each other in a chamber. ACP has a number of advantages over previously discussed solid parallel-plate press (SPP): (1) ACP uses gas or fluid layers to eliminate any direct contact between the solid plates and samples (mold and/or substrate), and, hence, removes any effects related to the imperfection of the solid plates. (2) Because the pressurized gas is conformal to the mold and substrate, regardless of their backside shapes or any dust particles on the backside, the pressure will be uniform everywhere over the entire imprint area; this uniformity is shown in Figure 6.7. (3) Isotropically applied gas pressure eliminates lateral shifting or rotation between the mold and substrate, reducing damage to the mold and prolonging mold lifetime. (4) ACP keeps the pressure on the mold and substrate at a preset value rather than the total force as in SSP, thus eliminating the “hot” spots (local high-pressure regions caused by small contact areas under a constant force) in SSP that damage the mold and the substrate. (5) Because a pressurized gas has much smaller thermal mass than a solid plate, when combined with radioactive direct heating to the samples and convection cooling, ACP shortens the thermal imprint time by orders of magnitude. For example, ACP can complete the nanoimprint process in seconds rather than in tens of minutes as in SPP.

6.2.4 Sequential Embossing/Imprinting Lithography (SEIL)

Inspired by the cold-drawing techniques developed for metallic compounds to compose microscale or nanoscale rods/wires, a new process, dubbed SEIL, was designed [15]. SEIL takes the easy-to-process nature of embossing/imprinting to create 750-nm pitch grating lines on soft substrates and then utilizes uniaxial stretching along the longitudinal direction of grating features to achieve a significant pitch reduction to 120 nm.

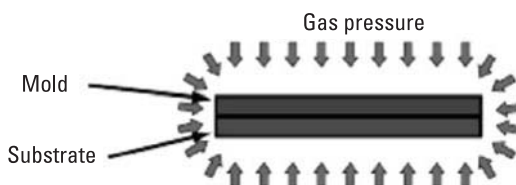


Figure 6.6 Schematic of the ACP nanoimprint principle. (Reprinted with permission from [14]. Copyright 2006 American Chemical Society.)

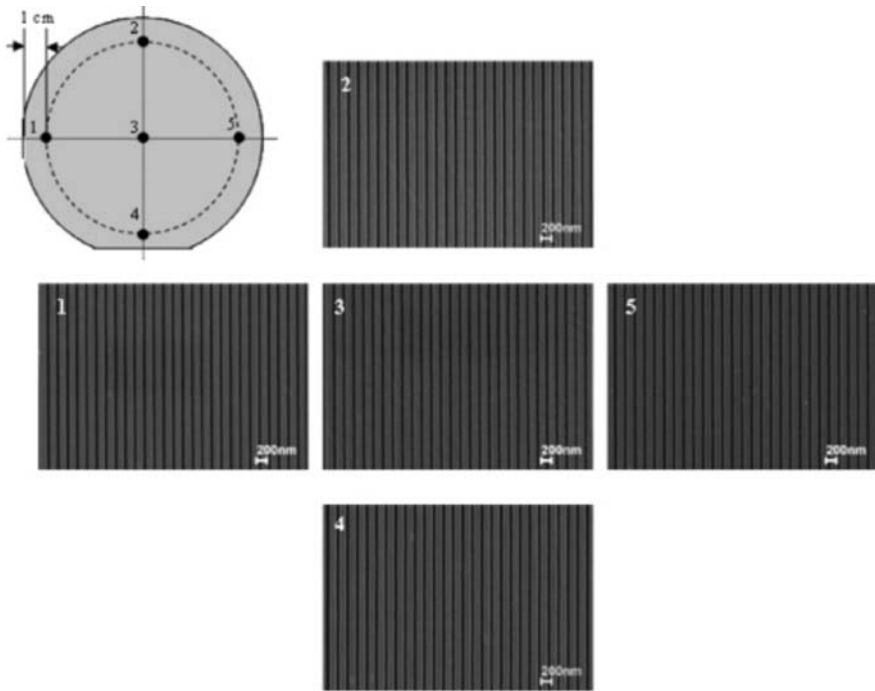


Figure 6.7 SEM images taken from five locations on the 100-mm-diameter sample imprinted using thermal NIL, one at the center and four on a circle 10 mm from the wafer edge. Each SEM is labeled with a number from 1 to 5, corresponding to the location from which it was taken. Uniform 200-nm period gratings are observed at all locations. (Reprinted with permission from [14]. Copyright 2006 American Chemical Society.)

In the process shown in Figure 6.8, a master pattern is imprinted into an elastomer containing a film of uncured elastomer. The elastomer is cured and then elongated to increase feature density and reduce feature size. Replication of this substrate yields a new master that can be used in further reduction steps. Compared to the previously demonstrated contraction approach, SEIL results in faithful pattern miniaturization in all aspects and, as a result, provides very effective control on density and dimension regulations.

6.3 AFM Lithography

Among the nanofabrication techniques available, scanning probe lithography (SPL) is another highly promising approach at the nanometer scale. In particular, extensive work has been carried out to explore the potential of nanolithographic

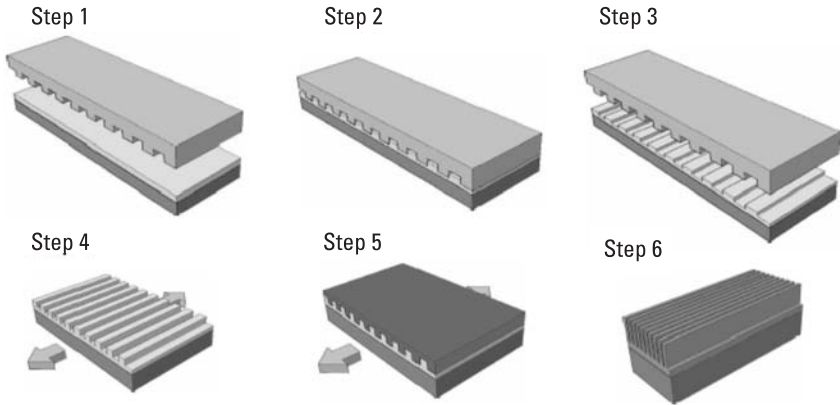


Figure 6.8 Process flow for sequential embossing/imprinting lithography (SEIL): (1) Coat a preelastomer (liquid) over a soft substrate; (2) emboss/imprint using a rigid mold to displace liquid materials underneath the mold; (3) solidify the liquid precursor after baking and remove the mold; (4) stretch the soft pattern to regulate feature dimension and density; (5) replicate the topographical features to a hard material; and (6) repeat steps (1) through (5) until the desired dimension or density is reached.

writing using sharp probe tips with an atomic force microscope (AFM). The advantages of AFM lithography are several-fold. First, it is simple, can be carried out under ambient conditions, and does not require sophisticated machinery. Second, AFM lithography can enable nanofabrication with feature sizes at or below the 10-nm level. Third, an AFM is capable of high-resolution, real-space imaging and can ensure alignment accuracy at the nanometer level for overlayers. The speed and throughput of AFM have also experienced significant improvements over the years with parallel operations of probe-tip arrays.

6.3.1 Scratching and Nanoindentation

AFM has been used for nanoscale material removal through both mechanical and chemical means. In the former, the materials are mechanically removed by direct tip scratching or plowing, while in the latter, the materials are removed by tip-induced electrochemical etching. By applying a certain amount of force on the tip or by controlling the cantilever deflection during scanning, the AFM tip has been successfully employed to “scratch” several hard surfaces of metals, oxides, and semiconductors, producing furrows with widths of tens of nanometers and depths of a few nanometers. Magno and Bennett [16] have used a Si_3N_4 cantilever tip in a contact mode to dig three 500-nm-long grooves on a two-layer GaSb/InAs film, as shown in Figure 6.9. The stiffness of the cantilever

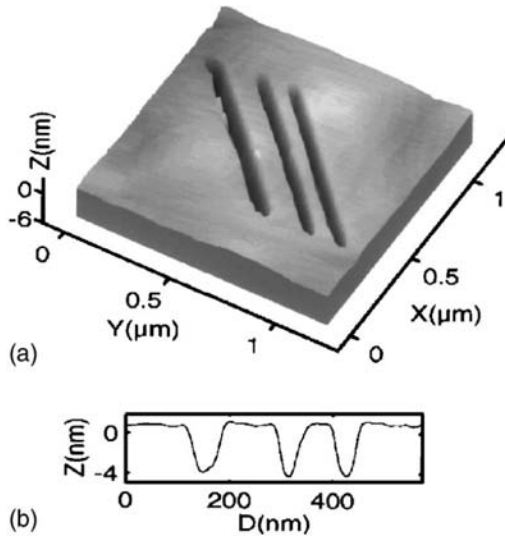


Figure 6.9 Three 500-nm-long grooves scribed on 5-nm GaSb/20-nm InAs substrate: (a) AFM image and (b) profile of grooves, 65 nm in width and 5 nm in depth, indicating that the deepest groove is just to penetrate through the GaSb into the InAs. (Reprinted with permission from [16]. Copyright 1997 American Institute of Physics.)

is 0.37 N/m and the applied forces vary from 100 to 150 nN with a scanning rate of 0.02 mm/s. The grooves produced on the 5-nm GaSb/20-nm InAs substrate are about 65 nm in width and 4.7 to 5.0 nm in depth. Because the GaSb layer is 5 nm thick, the deepest groove just penetrates through the GaSb layer into the InAs layer. For a conventional modulation-doped GaAs/AlGaAs heterostructure, the resulting grooves can form an arrangement of barriers in the electron layer. A type of heterostructure with a compensating *p*-type doped cap layer shows an electron enhancement if the cap layer is selectively removed. Etching a groove in these structures enables one to induce a 1D electron system. Both types of structures can be used to fabricate various ballistic quantum devices and Coulomb-blockade structures.

One of the major limitations in direct mechanical modification can be its shallow scratch depth because the applied tip forces cannot be too large due to possible damage of the tip or a decreased tip lifetime. Hu et al. [17] developed a bilayer resist system consisting of a 3-nm-thick Ti film on top of a 65-nm-thick PMMA layer. For patterning, the Si tip only needs to scratch the 3-nm Ti film with a tip force of a few micronewtons and the scratched pattern is then transferred to the PMMA layer underneath by an isotropic O₂ RIE. A liftoff procedure is applied to use the transferred PMMA pattern to make Si nanowires that are 30 nm in width and 20 nm in thickness and Cr nanodots of 20 to 40

nm in diameter and 20 nm in height. Note that directly scratching the PMMA may damage the substrate underneath if the applied scratch force is too great.

AFM has better reproducibility when applied to making thermally induced surface modifications [18]. Input power of 35 mW was used to heat the AFM tip to temperatures of up to 170°C for 4 ms to produce an array of sub-100-nm pits on polycarbonate substrate, as shown in Figure 6.10(a). IBM has extended this thermally induced modification concept to data storage applications. The extended concept, also called Millipede, uses an array of AFM-type probes to write, read, and erase data on very thin polymer films. Figures 6.10(b, c) illustrate the Millipede concept, which exploits the parallel operation of very large 2D, for example, 34×34 AFM cantilever arrays with an integrated tip for write/read/erase functionality. The latest report on Millipede has indicated employment of a 15-nm bit pitch, which enables a storage density of more than 1 Tbit/in².

6.3.2 Nanografting

A technique called *nanografting* has recently been developed by Liu's group as first reported in 1997 [19]. In that report, the authors combined AFM with the surface chemistry of thiol adsorption on gold surfaces. The procedure of nanografting is relatively simple, starting from imaging (under a small force) an alkanethiol self-assembled monolayer (SAM) in a liquid medium containing a different kind of thiol [Figure 6.11(a)]. As a higher force is applied during the scanning, the matrix thiol molecules are removed by the tip and transported into the solvent. Thiol molecules contained in the solution immediately adsorb onto the freshly exposed gold area following the scanning track of the AFM tip to form the designed nanostructures. The nanopatterns produced can then be characterized in situ by the same AFM tip at a reduced force.

Since its invention, many improvements and technological extensions of nanografting have been researched. Figure 6.11(b) shows an example in which

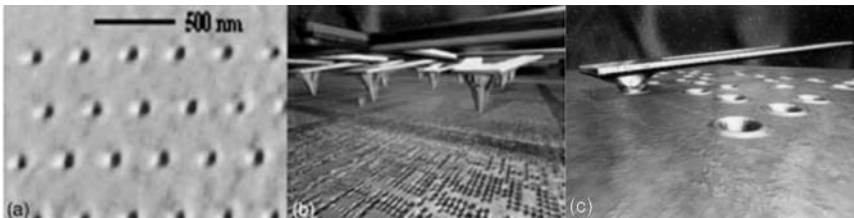


Figure 6.10 Thermomechanical writing by AFM: (a) AFM image of a sub-100-nm dot array written on polycarbonate using an electrically heated sharp-cantilever tip with 35-mW, 4-ms pulses and (b, c) schematics of IBM's Millipede. (Reprinted with permission from [18]. Copyright 1996 American Institute of Physics.)

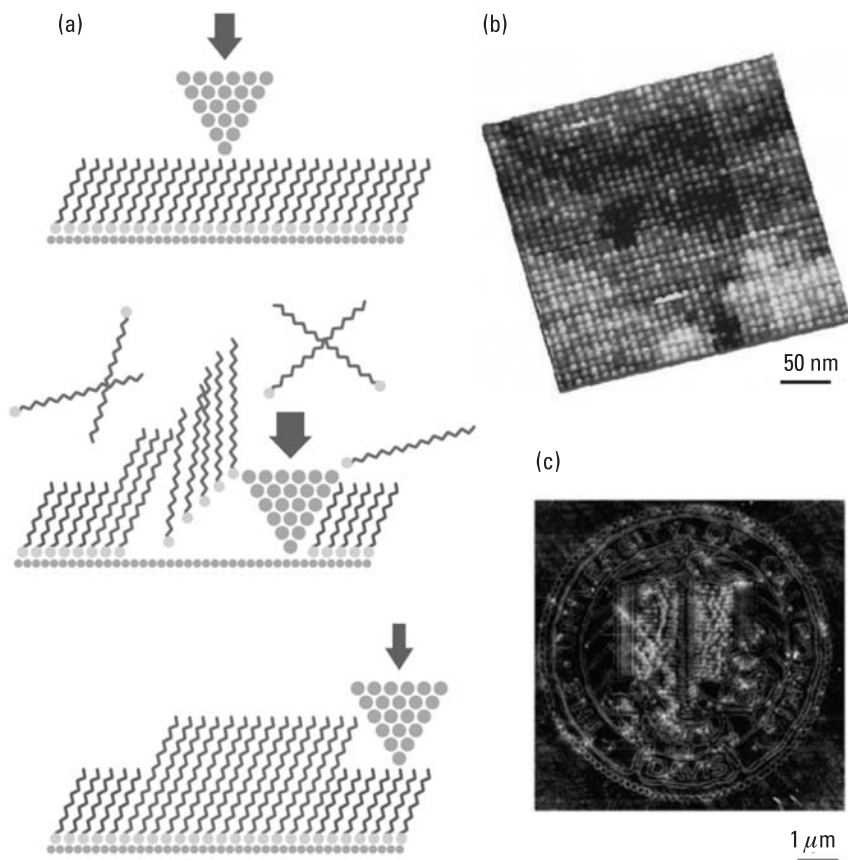


Figure 6.11 (a) Schematic showing the process of nanografting. (b) A 33×33 array of nanostructures of biotin-terminated thiols inlaid in hexanethiol matrix, produced in an ethanol medium using nanografting. (c) An $8 \mu\text{m} \times 8 \mu\text{m}$ lateral force image of the University of California at Davis seal produced by nanografting. (Reprinted with permission from the *Annual Review of Physical Chemistry*, Vol. 59. © 2008 Annual Reviews, [20] www.annualreviews.org.)

a 33×33 array of biotin-terminated thiols was produced on gold. Each element has a size of $5.2 \times 5.2 \text{ nm}^2$ with great uniformity. A *nanopen* reader and writer extends nanografting into ambient or nonreactive media. In a nanopen reader/writer, reactants are predeposited onto an AFM probe. The probe can then perform AFM imaging under low load and deposit the desired materials under high force on a gold surface by replacing the resist molecules.

Software and digitization in scanning probe lithography has also significantly improved automated nanolithography for the high-throughput produc-

tion of complex nanostructures and arrays. Figure 6.11(c) illustrates a revealing example in which the University of California at Davis's seal is fabricated by nanografting an aldehyde terminated thiol into a decanethiol SAM. The process took 10 min to complete and its finest width was 10 nm. Automated nanografting may be utilized in conjunction with multiple AFM tips in a 1D or 2D format for the parallel fabrication of nanopatterns.

Similar to many scanning probe lithography techniques, such as scanning tunneling microscopy-based lithography, dip-pen nanolithography, local oxidation nanolithography, and local chemical or electrochemical lithography, nanografting has many advantages in potential applications in materials science and the nanotechnology industry. For instance, its spatial resolution is high; thus the production and characterization of sub-100-nm structures become feasible. Nanografting also shows great promise in materials science by producing various functional nanostructures, including $-\text{OH}$, $-\text{COOH}$, $-\text{CHO}$, $-\text{NH}_2$, $-\text{NHS}$, biotin, $-\text{CF}_3$, carbohydrate, and nucleotides. One can construct 3D nanostructures using pattern transfer by further surface reactions. Moreover, the versatility of nanografting has been demonstrated by the creation of nanostructures of large molecules and biological molecules, such as nanoparticles of metals, DNA, ligands, and proteins. Many researchers and laboratories that have AFM capabilities also favor nanografting for its simplicity.

Four unique aspects of nanografting make it a new and powerful tool in surface physical chemistry [20]. First, by not requiring any tip modification and eliminating surface diffusion using a self-assembly monolayer as a resist, nanografting harnesses the highest spatial resolution that AFM can offer both in nanostructure production and in characterization. Second, nanografting can work in versatile chemical environments (e.g., wet chemistry), which enables direct mimicking and monitoring of real-surface reactions in situ, in real time, and with molecular resolution. Third, nanografting is an active tool for exploring the reaction mechanism, kinetics, and products. Fourth, multiplexing is enabled within one experiment by producing multicomponent nanostructures and then introducing designed reactants while monitoring the outcome in situ.

6.4 Polymer Pen Lithography (PPL)

Recently, a new method, called polymer pen lithography (PPL), combined with parallel replication and serial writing has been reported by Huo et al. [21]. It is a low-cost, cantilever-free lithographic approach that, thus far, allows a digitized pattern to be printed at spot sizes ranging from 90 nm to hundreds of micrometers simply by changing the force and time over which the ink is delivered. In contrast with dip-pen nanolithography (DPN) and other SPM-based lithographies, which

typically use hard silicon-based cantilevers, PPL uses elastomeric tips without cantilevers to deliver ink.

A typical polymer pen array (Figure 6.12) contains thousands of pyramid-shaped tips that are made with a master prepared by conventional photolithography and subsequent wet chemical etching (Figure 6.13). The pyramids are connected by a thin PDMS backing layer (50 to 100 μm thick) that is adhered to a glass support prior to curing. The glass support and thin backing layer significantly improve the uniformity of the polymer pen array over large areas; to date, up to an entire 3-in. (7.5-cm) wafer surface [Figure 6.12(b)]. When the sharp tips of the polymer pens are brought in contact with a substrate, ink is delivered at the points of contact [Figure 6.12(a)].

An important feature for ensuring uniform ink delivery over large areas is that the amount of light reflected from the tips increases greatly when the tips make contact with the substrate. Thus, the transparent elastomer polymer pen array allows us to determine visually when all of the pens are in contact with an underlying substrate. The otherwise daunting task of leveling the array can be tackled in an experimentally straightforward manner. All of the pens are

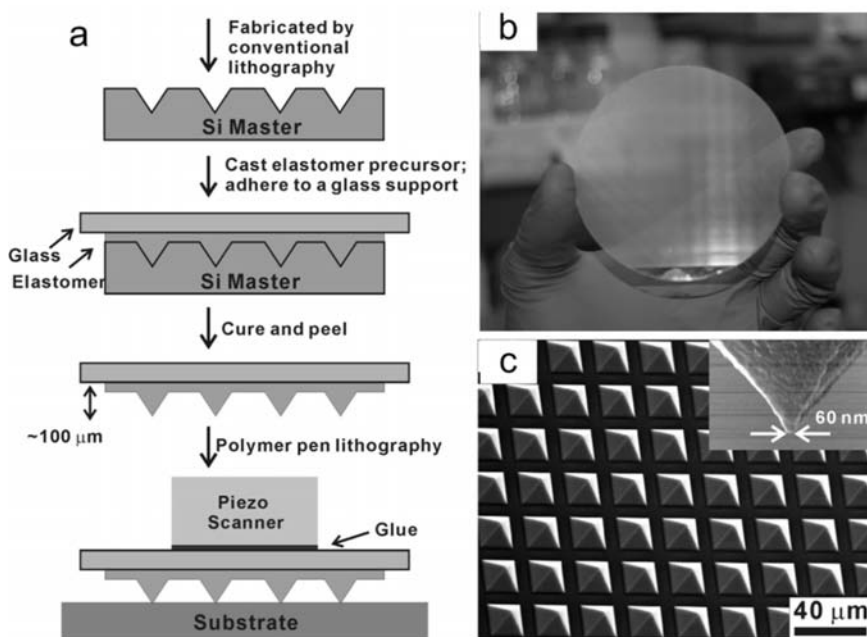


Figure 6.12 (a) A schematic illustration of the polymer pen lithography setup. (b) A photograph of an 11-million pen array. (c) SEM image of the polymer pen array. The average tip radius of curvature is 70 ± 10 nm (inset). (Reprinted with permission from [21]. Copyright 2008 American Association for the Advancement of Science.)

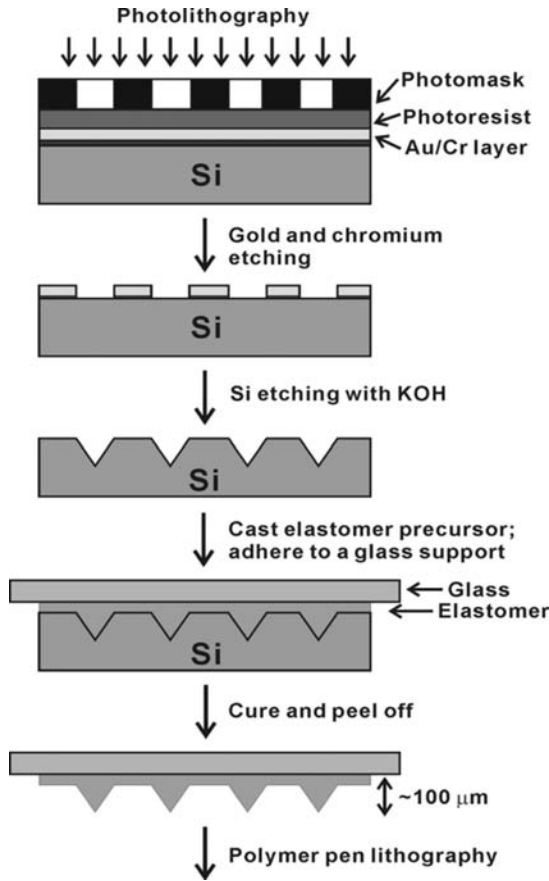


Figure 6.13 Schematic diagram of the polymer pen array fabrication process. (Reprinted with permission from [21]. Copyright 2008 American Association for the Advancement of Science.)

remarkably uniform in size and shape, with an average tip radius of $70 \pm 10 \text{ nm}$ [Figure 6.12(c)]. In principle, this value could be reduced substantially with higher quality masters and stiffer elastomers.

The force dependence and maskless nature of PPL allow many structural variants to be created without the hurdle of designing a new master via a throughput-impaired serial process. In addition, PPL can be used with sub-100-nm resolution with the registration capabilities of a closed-loop scanner. For example, in [21] PPL was used to generate 15,000 replicas of the 2008 Beijing Olympic logo on gold with MHA as the ink and subsequent wet chemical etching [Figure 6.14(a)]. Each logo was generated using the multiscale capabilities of PPL from a $70 \times 60\text{-}\mu\text{m}^2$ bitmap. The letters and numbers “Beijing 2008”

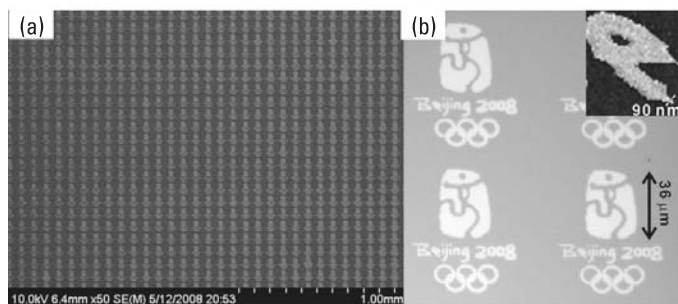


Figure 6.14 (a) SEM image of a representative region of approximately 15,000 miniaturized duplicates of the 2008 Beijing Olympic logo. (b) A zoom-in optical image of a representative replica. The inset shows a magnified SEM image of the letter “e.” (Reprinted with permission from [21]. Copyright 2008 American Association for the Advancement of Science.)

were generated from $\sim 20,000$ 90-nm dots (initial contact), and the picture and Olympic rings were made from $\sim 4,000$ 600-nm dots at higher array-substrate contact forces (relative piezo extension = $1 \mu\text{m}$). These structures were created by holding the pen array at each spot for 0.05 sec and traveling between spots at a speed of $60 \mu\text{m/s}$. A representative portion of the approximately 15,000 replicas (yield > 99%) generated across the 1-cm^2 substrate demonstrated the uniformity and stability of the process [Figure 6.14(b)]. Importantly, the total time required to fabricate all of these structures was less than 40 min.

6.5 Templated Self-Assembly of Block Copolymers

Materials that can self-assemble comprise the building blocks of future advanced nanotechnologies based on bottom-up fabrication methods [22]. Self-organizing materials provide simple and low-cost processes to make large-area periodic nanostructures. Alternatively, top-down lithographic approaches offer arbitrary geometrical designs and superior nanometer-level precision, accuracy, and registration. By combining bottom-up self-assembly with top-down patterned templates, templated self-assembly (TSA) can provide rich opportunities for fundamental studies of self-assembly behavior in confined environments, as well as a source of innovation in nanofabrication methods.

Self-assembly has been the focus of much research in the last four decades. These efforts have produced a solid foundation for understanding the physics and chemistry of self-organizing processes in bulk materials. Sophisticated lithographic tools for fabricating complex 2D structures have also been developed during this time. Recent efforts have striven to bring the two fields (self-assembly and lithography) together.

TSA is a method of eliminating defects and inducing registration and orientation in thin films of materials using artificial topographical and/or chemically patterned templates. In contrast to conventional epitaxy in which the lattice of a thin film bears a well-defined relationship to the lattice of the underlying substrate, templates for TSA are not required to be crystalline materials. In the concept of template self-assembly, the topography and/or chemical pattern of the top templates instead of the atomic lattice of the bottom substrate are used to guide the organization of the component materials. The characteristic feature size L_S of templates ranges from the characteristic length scale, L_0 , of the self-assembled materials to sizes much larger than L_0 , as illustrated in Figure 6.15.

Yin et al. [23] demonstrated that colloidal arrays deposited with a flow technique have generated ordered structures in commensurate templates, and different, but still ordered, structures in incommensurate templates. Figure 6.16 outlines the schematic procedure that was used by Yin et al. [23]. The monodispersed polymer beads were first self-assembled into complex aggregates under the physical confinement exerted by the holes patterned in a thin film of photoresist spin coated on the surface of a glass substrate [Figure 6.16(b)]. Next an aqueous dispersion of the polymer beads was added into this cell consisting of two glass substrates [Figure 6.16(a)], and the rear front of this confined slug of liquid was allowed to move slowly along the direction indicated by the arrow by removing the solvent from the front edge of the liquid layer. The patterned 2D array of cylindrical holes etched in the surface of the bottom substrate simply served as physical traps for the liquid and the polymer beads.

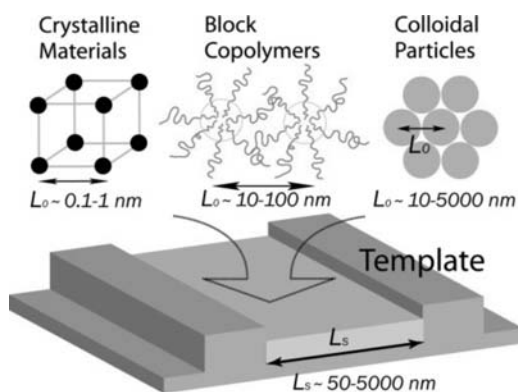


Figure 6.15 Illustration of some types of TSA systems. Characteristic lengths (L_0) of crystalline materials, block copolymers, and colloid assemblies and the characteristic length (L_S) of the template are indicated. (Reprinted with permission from [22]. Copyright 2006 Wiley InterScience.)

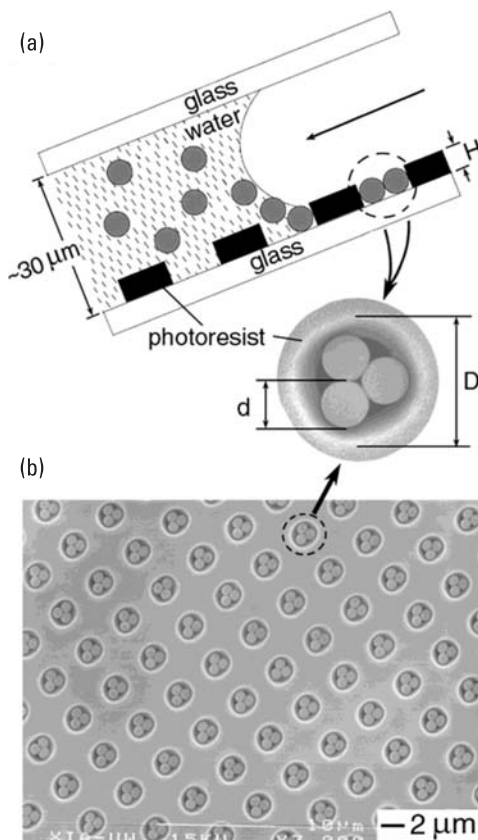












Figure 6.16 (a, b) A schematic illustration of the procedure that we used to assemble spherical colloids into well-controlled clusters under the physical confinement exerted by the 2D array of cylindrical holes patterned in a thin film of photoresist. The size, shape, and structure of these clusters could all be easily controlled by changing the ratios between the dimensions (H and D) of the holes and the diameter (d) of the polymer beads. (Reprinted with permission from [23]. Copyright 2001 Wiley InterScience.)

As the rear front of the liquid moved, the capillary forces exerted on this interface dragged the polymer beads across the surface of the bottom substrate until they were physically trapped by the holes. The maximum number of polymer beads that could possibly be retained in each hole and thus the structural arrangement among these beads were solely determined by the ratios between the dimensions (D and H) of the holes and the diameter (d) of the polymer beads. A range of uniform, polygonal (or polyhedral) clusters could be formed in such 2D arrays of holes by simply controlling these ratios (Table 6.1). The difference in value between any two adjacent entries in Table 6.1 fully

Table 6.1

Control over the Structure of the Cluster by Changing the Ratio Between the Dimensions (D , H) of the Holes and the Diameter (d) of the Polymer Beads

D/d	Single-Layered ($0.5d \leq H \leq 1.37d$)	Double-Layered ($1.37d \leq H \leq 2.23d$)
1.00–2.00	monomer 	dimer 
2.00–2.15	dimer 	tetrahedron 
2.15–2.41	triangle 	octahedron 
2.41–2.70	square 	bi-square-pyramid 
2.70–3.00	pentagon 	
3.00–3.30	hexagon 	

Source: [23].

demonstrates the level of control that can be easily achieved. When the solvent was completely evaporated, those polymer beads trapped in each hole always tended toward physical contact as a result of the attractive capillary forces among them.

The yield of the desired cluster (as calculated from the geometric parameters) of this process could be as high as 95% [see Figure 6.16(b)]. Commonly

observed defects were clusters with fewer polymer beads than one would expect. Such defects could be eliminated by adjusting the concentration of the dispersion of polymer beads, by flowing a relatively dilute dispersion across the surface more than one time, or by switching to heavier particles made of inorganic materials. Defect-free, 2D arrays of clusters over areas as large as $2 \times 2 \text{ mm}^2$ (or as many as ~ 105 elements in the array) was able to be obtained.

References

- [1] Xia, Y. N., and G. M. Whitesides, "Soft Lithography," *Annual Review of Materials Science*, Vol. 28, 1998, pp. 153–184.
- [2] Liu, G. Y., S. Xu, and Y. L. Qian, "Nanofabrication of Self-Assembled Monolayers Using Scanning Probe Lithography," *Accounts of Chemical Research*, Vol. 33, No. 7, 2000, pp. 457–466.
- [3] Sotomayor Torres, C. M., *Alternative Lithography: Unleashing the Potential of Nanotechnology*, New York: Kluwer Academic/Plenum Publishers, 2003.
- [4] Schift, H., "Nanoimprint Lithography: An Old Story in Modern Times? A Review," *J. of Vacuum Science & Technology B*, Vol. 26, No. 2, 2008, pp. 458–480.
- [5] Ito, T., and S. Okazali, "Pushing the Limits of Lithography," *Nature*, Vol. 406, 2000, pp. 1027–1031.
- [6] Kumar, A., and G. M. Whitesides, "Features of Gold Having Micrometer to Centimeter Dimensions Can Be Formed Through a Combination of Stamping with an Elastomeric Stamp and an Alkanethiol 'Ink' Followed by Chemical Etching," *Applied Physics Letters*, Vol. 63, No. 14, 1993, pp. 2002–2004.
- [7] Chou, S. Y., P. R. Krauss, and P. J. Renstrom, "Imprint of Sub-25 nm Vias and Trenches in Polymers," *Applied Physics Letters*, Vol. 67, No. 21, 1995, pp. 3114–3116.
- [8] Xu, S., et al., "Fabrication of Nanometer Scale Patterns Within Self-Assembled Monolayers by Nanografting," *Langmuir*, Vol. 15, No. 21, 1999, pp. 7244–7251.
- [9] Maoz, R., S. R. Cohen, and J. Sagiv, "Nanoelectrochemical Patterning of Monolayer Surfaces: Toward Spatially Defined Self-Assembly of Nanostructures," *Advanced Materials*, Vol. 11, No. 1, 1999, pp. 55–61.
- [10] Kumar, A., et al., "The Use of Self-Assembled Monolayers and a Selective Etch to Generate Patterned Gold Features," *J. of the American Chemical Society*, Vol. 114, No. 23, 1992, pp. 9188–9189.
- [11] Austin, M. D., et al., "6 nm Half-Pitch Lines and $0.04 \mu\text{m}^2$ Static Random Access Memory Patterns by Nanoimprint Lithography," *Nanotechnology*, Vol. 16, No. 8, 2005, pp. 1058–1061.
- [12] Chou, S. Y., et al., "Sub-10 nm Imprint Lithography and Applications," *J. of Vacuum Science & Technology B*, Vol. 15, No. 6, 1997, pp. 2897–2904.
- [13] Li, M. T., L. Chen, and S. Y. Chou, "Direct Three-Dimensional Patterning Using Nanoimprint Lithography," *Applied Physics Letters*, Vol. 78, No. 21, 2001, pp. 3322–3324.

-
- [14] Gao, H., et al., "Air Cushion Press for Excellent Uniformity, High Yield, and Fast Nanoimprint Across a 100 mm Field," *Nano Letters*, Vol. 6, No. 11, 2006, pp. 2438–2441.
 - [15] Lin, H. J., et al., "Collective Buckling of Periodic Soft Nanostructures on Surfaces and Promotion for Nanolithography," *J. of Physics Chemistry C*, Vol. 111, 2007, pp. 13348–13356.
 - [16] Magno, R., and B. R. Bennett, "Nanostructure Patterns Written in III–V Semiconductors by an Atomic Force Microscope," *Applied Physics Letters*, Vol. 70, No. 14, 1997, pp. 1855–1857.
 - [17] Hu, S., et al., "Fabrication of Silicon and Metal Nanowires and Dots Using Mechanical Atomic Force Lithography," *J. of Vacuum Science & Technology B*, Vol. 16, No. 5, 1998, pp. 2822–2824.
 - [18] Mamin, H. J., "Thermal Writing Using a Heated Atomic Force Microscope Tip," *Applied Physics Letters*, Vol. 69, No. 3, 1996, pp. 433–435.
 - [19] Xu, S., and G. Y. Liu, "Nanometer-Scale Fabrication by Simultaneous Nanoshaving and Molecular Self-Assembly," *Langmuir*, Vol. 13, No. 2, 1997, pp. 127–129.
 - [20] Liu, M. Z., N. A. Amro, and G. Y. Liu, "Nanografting for Surface Physical Chemistry," *Annual Review of Physical Chemistry*, Vol. 59, 2008, pp. 367–386.
 - [21] Huo, F. W., et al., "Polymer Pen Lithography," *Science*, Vol. 321, 2008, pp. 1658–1660.
 - [22] Cheng, J. Y., et al., "Templated Self-Assembly of Block Copolymers: Top-Down Helps Bottom-Up," *Advanced Materials*, Vol. 18, No. 19, 2006, pp. 2505–2521.
 - [23] Yin, Y. D., and Y. N. Xia, "Self-Assembly of Monodispersed Spherical Colloids into Complex Aggregates with Well-Defined Sizes, Shapes, and Structures," *Advanced Materials*, Vol. 13, No. 4, 2001, pp. 267–271.

7

Synthesis of Nanoparticles and Their Self-Assembly: Bottom-Up Approach

Synthesis of nanoscale spheres or nanoparticles (NPs) and assembly of them into well-defined macroscopic structures is a promising strategy for “bottom-up” materials design. During the past two decades, NPs have attracted serious attention worldwide. Remarkable achievements in innovative synthetic routes and growth mechanisms have been made. Prepared NPs have a wide variety of morphologies, including cubes, polyhedron, rods, rings, wires, dots, and many other exotic shapes. The sizes of these NPs are usually confined to less than 100 nm, at least in one dimension. Such small sizes confer on NPs unique properties such as large surface area, quantum effect, surface plasmon resonance, and superparamagnetism that are frequently different from the properties of bulk materials, making NPs ideal candidates for applications in single-electron devices, nanoelectronics, and so forth. After the synthetic approaches have matured, the use of NPs as building blocks for multidimensional assemblies with controlled sizes and morphologies becomes possible. Nanoparticles can be organized into a variety of predetermined structures via noncovalent interactions, such as dipolar, electrostatic, and van der Waals interactions.

In this chapter, we discuss the mechanisms for synthesizing and growing NPs. Subsequently, the self-assembly of NPs into different nanoscale or macroscopic structures is presented.

7.1 Synthesis of Nanoparticles

Assembly of nanoparticles into highly ordered architectures requires NPs with a uniform shape and a monodispersed size. Many well-studied and controlled

synthetic protocols for preparing uniform NPs have been established, such as coprecipitation, sol-gel processing, microemulsions, hydrothermal/solvothermal methods, template syntheses, biomimetic syntheses, supercritical fluid synthesis, and ionic liquid synthesis. In this section, we focus on wet-chemical methods.

7.1.1 Coprecipitation

Among all of the synthetic strategies, the coprecipitation method is the most used. In the “old days,” NPs were made by coprecipitation of solids from aqueous solutions followed by thermal decomposition of those precipitates. The coprecipitation process has several steps, including nucleation, growth, coarsening, and/or agglomeration. To prepare monodispersed NPs, the nucleation and growth stages must be under precise control. This was achieved by adjusting the reaction parameters such as by altering the mixing ratio of the reactants, varying the injection sequences, adding surfactants, and changing solvents. Among these parameters, an appropriate reaction rate plays a key role. Typically, rapid nucleation and relatively slow growth afterward will lead to the formation of monodispersed NPs.

In the slow growth mode, molecules of NPs generally follow the Oswald ripening process, whereas the rapid growth mode will always lead to an irregular morphology and scattered size distribution for the NPs. The Oswald ripening phenomenon can be briefly described as follows: Small particles vanish while large particles continue to grow in size. It stems from the fact that molecules on the surface of a particle are energetically less stable than the ones already well ordered and packed in the interior. Large particles, with their lower surface-to-volume ratio, have a lower energy state. As the system tries to lower its overall energy, molecules on the surface of a small particle will tend to diffuse in the solution and add to the surface of a larger particle. For details on this mechanism and a few examples of synthesis, we suggest readers read a full review article by B. L. Cushing et al. [1].

Using aqueous solutions to synthesize semiconductor NPs is feasible and also environmentally friendly. For instance, NiO, ZnO, and SnO₂ NPs were obtained after being precipitated from an aqueous solution of metal chloride [2]. This reaction usually gives an amorphous product; subsequent calcination or annealing is necessary in order to render the NPs crystalline, in which some agglomeration will be unavoidable. In rare cases, crystalline NPs can be precipitated from aqueous solutions. For instance, 50- to 60-nm rutile-structured TiO₂ was prepared by precipitating TiCl₃ within NH₄OH. In this process the NPs were stabilized by poly(methylmethacrylate) [3]. However, many parameters need to be precisely controlled in aqueous media, for instance, pH level, method of mixing, and the nature and the concentration of anions. Often, the complexity

of the aqueous approach leads to a situation in which slight changes in experimental conditions could have a strong influence on particle morphology [4].

A nonaqueous approach through the use of an organic solvent is able to overcome some of these issues of aqueous solutions. Although these strategies are not completely new and some of the basic principles were investigated a century ago, only recently have they been developed in a more systematic fashion. Overall, a nonaqueous approach provides us with ample opportunities to create inorganic semiconductor NPs including metal oxides and metal chalcogenides with controllable morphologies and size distributions. For instance, Sun et al. [5] prepared Fe_3O_4 crystals with a diameter of 4 nm by dissolving $\text{Fe}(\text{acac})_3$, 1,2-hexadecanediol, oleic acid, and oleylamine in diphenyl ether before refluxing. Rockenberger et al. [6] developed a method for preparing surfactant-capped oxide NPs by decomposition of metal-Cupferron complexes, M^xCup_x ($\text{Cup} = \text{C}_6\text{H}_5\text{N}(\text{NO})\text{O}^-$). These NPs generally have a diameter between 4 and 10 nm and can be easily dispersed in toluene [6].

NPs of chalcogenide are important semiconducting materials because they offer controllable bandgaps. These gaps are regulated through variations in the size of the NPs, which confer chalcogenide NPs large potential in quantum dot applications. Classic metathesis reactions between metal and chalcogenide ions usually produce agglomerated crystallines with widely disparate sizes and shapes. In the 1990s, good strategies for preparing a variety of nanoscale II-VI and III-V semiconductors were established by using organic chalcogenide, metal molecular carriers, and organic coordinating solvents as the reaction medium. This type of synthesis is based on the pyrolysis of organometallic reagents by injection into a hot coordinating solvent. It provides temporally discrete nucleation and permits the controlled growth of macroscopic quantities of nanocrystallines [7, 8]. For instance, Murray et al. [8] reported on the preparation of CdSe NPs from organic metallic precursors. Further work from other groups revealed success in synthesizing other shapes for CdSe NPs (Figure 7.1). Formation of these morphologies was explained on the basis of an effective monomer model and selective adsorption of surfactants on different crystallographic faces [7].

To prepare anisotropic semiconductor NPs, multiple surfactants are usually required. However, some literature has reported on the use of a single-surfactant system to prepare anisotropic NPs [11, 12]. Cozzolli et al. [12] reported on the synthesis of ZnSe NPs using a monosurfactant amine system; the exact mechanism responsible for the successful shape evolution was not fully explained. In contrast to the aqueous system, which is sensitive to experimental conditions, nonaqueous strategies are much more robust. As a consequence, most of the processes are highly reproducible and easy to scale up to gram quantities. The nonaqueous routes also have disadvantages; for instance, they require a long period of time (from several days to weeks) to complete the precipitate process and toxic solvents are used in the process.

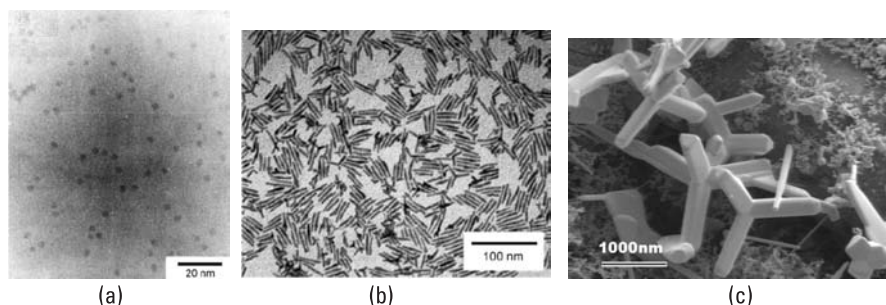


Figure 7.1 (a–c) From left to right, TEM micrographs of CdSe nanodots [8] and nanorods [9] and an SEM image of tetrapods [10]. [(a) Reprinted with permission from [8]. Copyright 1993 American Chemical Society. (b) Reprinted with permission from [9]. Copyright 2000 American Chemical Society. (c) Reprinted with permission from [10]. Copyright 2006 Wiley Interscience.]

7.1.2 Sol-Gel Process

The sol-gel process can be defined as the hydrolysis and condensation of a liquid precursor to a solid. It was reviewed by Hench and West in 1990 [13] and also by Brinker and Scherer [14]. Typically, the starting precursors can be either an inorganic salt or organic species such as a metal alkoxide. The whole process can be characterized by several distinct steps (Figure 7.2): formation of stable solutions of precursors (the sol); further reaction of the sol with a bridged, rigid, porous network (the gel) enclosing a continuous liquid phase by gelation; drying via the removal of liquids from the gel network; and densification and decomposition of the gels at high temperature. The nature of the specific reactions involved in hydrolysis and condensation differ substantially between various types of possible precursors. Finding a suitable precursor and solvent is the key to the synthesis of monodispersed semiconductor NPs by means of sol-gel processes.

In an aqueous sol-gel system, a four-coordinated tetraethoxy ($M(\text{OEt})_4$) is a widely used metallic precursor. The partial charge, δ , on metal of the $M(\text{OEt})_4$ is high, +0.32 for Si, +0.63 for Ti, and +0.74 for Zr, which causes a rapid rate of hydrolysis. To prevent such an undesirable situation, one or more hydroxylic ligands could be allowed to bind with the metal cores. When a transitional metal M of charge $Z+$ is dispersed in an aqueous solution, water molecules can coordinate with the metal core by transferring electrons from their bonding orbitals to empty d orbitals of metals by forming $[M(\text{OH}_2)_h]^{Z+}$ ion, where h is the coordination number of the metallic cation. Afterward, pH-dependent hydrolysis will take place to yield $[M=\text{O}]^{(Z-2)+}$. When this reaction occurs too fast, controlling the hydrolysis rate may require the selection of a nonaqueous solvent. The hydrolysis can then be controlled by regulating

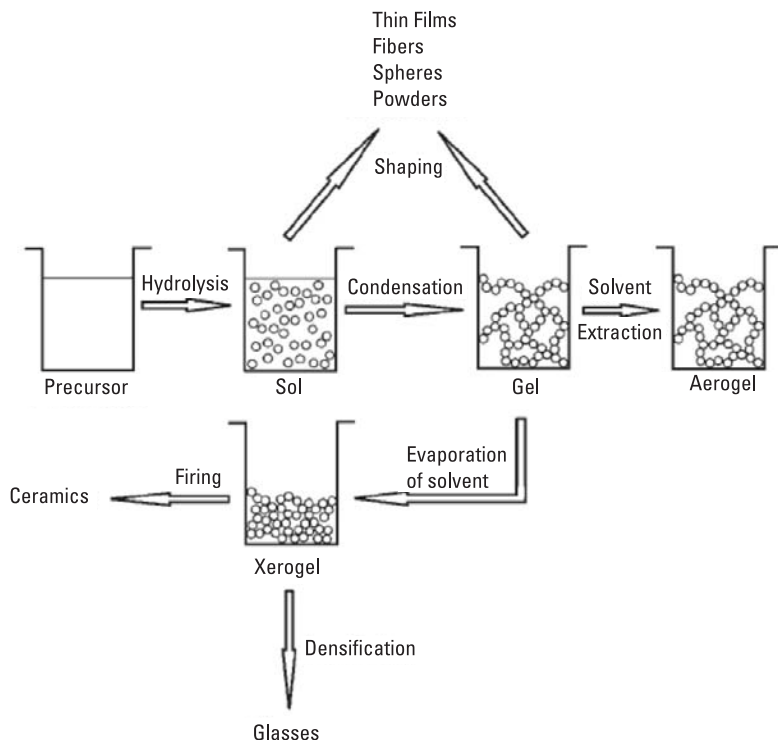


Figure 7.2 Typical steps in the sol-gel process [4]. (Reprinted with permission from [4]. Copyright 2006 Wiley Interscience.)

the water amount, instead of the pH, in an aqueous condition. Researchers have reported at length on the synthesis of semiconductor NPs by means of a nonaqueous sol-gel approach [15], for instance, TiO_2 [16, 17], FeO_x [18, 19], MnO [20, 21], CeO_2 [22], Co_3O_4 [23], NiO [24, 25], ZrO_2 [26, 27], InO_2 [28], SnO_2 [29], and WO_x [30]. A wide variety of metal-oxide precursors has been used through these procedures, ranging from organometallic complexes to metal halides, alkoxides, and acetylacetonates.

In most cases, a common feature shared by these precursors is the use of surfactants and coordinating solvents as stabilizing ligands. The use of ligands can often lead to the NPs being able to resist agglomeration between them. Ligands also promote dispersibility in solvents. Advantageously, these surface ligands can be easily exchanged with other families of ligands, offering control over the surface properties. However, some literature demonstrates that the surfactants can influence the toxicity of the NPs [31, 32]. On the other hand, in a nonaqueous approach, the oxygen in coordination with a metal cation is

provided by donors, such as ethers, alcohols, or alkoxides. Formation of an M-O-M linkage occurs through the condensation reaction between the ligands coordinated between two metal centers followed by the elimination of the organic fragments. There are several types of elimination, including alkyl halide elimination [see (7.1)], ether elimination [see (7.2)], ester elimination [see (7.3)], and other mechanisms (C-C bond formation, aldol condensation, and thermal decomposition) that are not discussed here:



The broad variety and low cost of metal halides make them popular candidates as precursors in the synthesis of metal-oxide gels. These processes involve condensation between metal alkoxides and metal halides under the release of an alkyl halide [see (7.1)]. The first synthesis of crystalline anatase nanoparticles based on alkyl halide elimination was reported by Trentler et al. in 1999 [33] using trioctylphosphine oxide (TOPO) as a surfactant to yield phase-pure materials. This strategy was later applied to the synthesis of Zirconia and other semiconductor metal oxides. Figure 7.3(a) is a high-resolution transmission electron microscopic (HRTEM) image of 4.0-nm Zirconia NPs obtained by this approach [34]. When zirconium bromide was used as a precursor instead of zirconium chloride, monodispersed NPs with an average size of 2.9 nm were yielded. However, one has to keep in mind that the NPs formed in the reaction mixture may even be able to catalyze undesired side reactions, such as ether elimination, making it impossible to distinguish between the direct products and side reaction products. Ether elimination pathways were developed much earlier than the alkyl halide ones to prepare oxides in 1986 [35]. Recently, Pinna et al. [36] reported the synthesis of HfO_2 and Ta_2O_5 NPs via ether elimination [Figures 7.3(b, c)]. In ester elimination routes, the first example used TiCl_4 as a precursor and added it to an alcohol/acid mixture followed by solvothermal treatment, which lead to highly crystalline anatase at rather mild conditions [37]. Quite analogous to ester eliminations are amide eliminations. Reacting metal oleates with amines can induce the controlled growth of oxides such as titania nanorods [Figure 7.3(d)] [16].

The synthesis of metal-oxide NPs in organic solvents by sol-gel routes opens the possibility of extending reaction principles from general organic chemistry to NP synthesis. However, many of the details of these extensions, which lead to the preparation of a certain NP with a desired composition, structure, size distribution, morphology, and even properties, are not yet fully understood, making the task a challenging one.

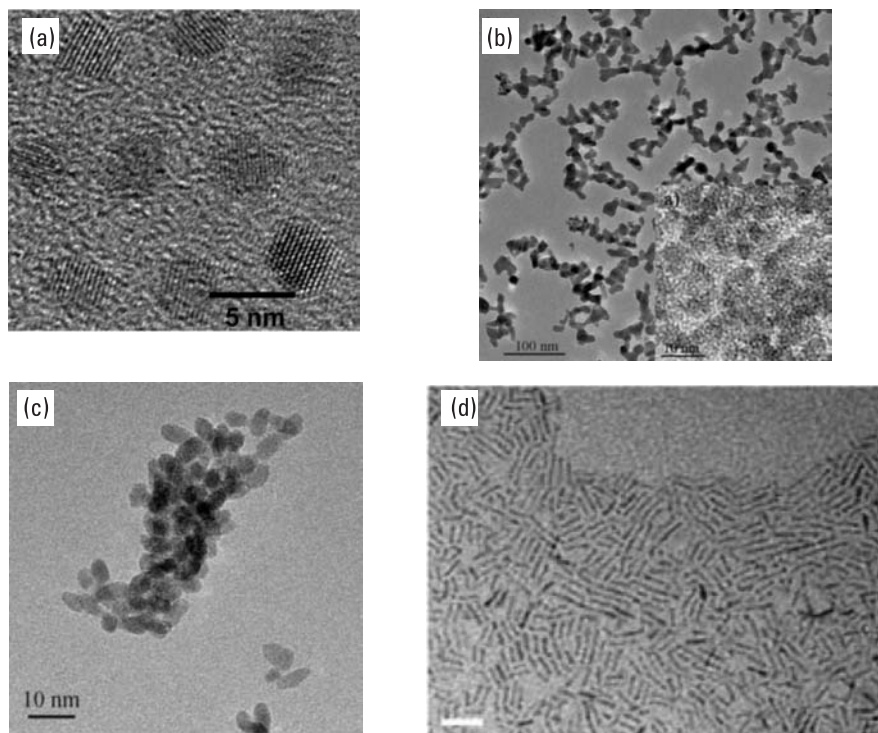


Figure 7.3 TEM images of (a) 5-nm ZnO_2 NPs [34], (b) 5-nm Ta_2O_5 NPs [36], (c) 6-nm length and 4-nm width HfO_2 NPs [36] and, (d) 3.8-nm length TiO_2 nanorods [16]. [(a) Reprinted with permission from [34]. Copyright 2003 American Chemical Society. (b, c) Reprinted with permission from [34]. Copyright 2004 Wiley Interscience. (d) Reprinted with permission from [16]. Copyright 2005 Wiley Interscience.]

7.1.3 Microemulsions

Early in 1943, Hoar and Schulman reported that certain combinations of water, oil, surfactant, and an alcohol or amine-based cosurfactant produced clear and homogeneous solutions, which they named *microemulsions* [38]. When the surfactant (usually with a hydrophilic head and a hydrophobic tail) is added to a mixture of water and oil (e.g., long-chain hydrocarbons), spherical aggregates are formed, in which polar ends of surfactant molecules orient toward the center. This model is visually demonstrated in Figure 7.4.

In contrast to ordinary emulsions, microemulsions form upon simple mixing of the components and do not require the high shear conditions generally used in the formation of ordinary emulsions. There are two basic types of microemulsions: direct microemulsion (oil dispersed in water, o/w) and the reverse (water dispersed in oil, w/o). When two immiscible phases (water and

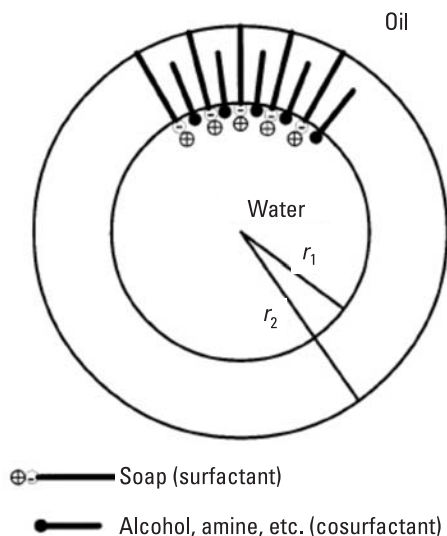


Figure 7.4 Reverse micelle model [1]. (Reprinted with permission from [1]. Copyright 2004 American Chemical Society.)

“oil”) are present, surfactant molecules form a monolayer along the interface between the oil and water, with the hydrophobic tails of the surfactant molecules dissolved in the oil phase and the hydrophilic head groups in the aqueous phase. In a binary system (water/surfactant or oil/surfactant), self-assembled nanostructures can be formed, ranging from (inverted) spherical and cylindrical micelles to lamellar phases and bicontinuous microemulsions, which may coexist with predominantly oil or aqueous phases. The spherical micelles can serve as small reactors that are loaded with metal salts; the addition of organically soluble chalcogenide reagents could result in the formation of semiconductor NPs [39, 40]. The size of the micelle does not physically affect the size of the particles, rather it affects the spatial distribution of the reagents. Shape anisotropy in semiconductor NPs was achieved by using cosurfactants, such as the anionic bis(2-ethylhexyl)sulfosuccinate (AOT) and the zwitterionic phospholipid *L*- α -phosphatidylcholine (lecithin) [41]. Using a similar approach, triangular CdS nanocrystals were synthesized [40]. Although interesting, the sizes of the NPs produced by microemulsion routes are rather restricted.

7.1.4 Hydrothermal/Solvothermal Methods

In a sealed vessel, solvent can be brought to a temperature well above its boiling point. Chemical reactions taking place under such a condition are referred as

undergoing a *solvothermal process*. When the temperature and pressure are above certain values (critical point), this solvothermal process is also called a *supercritical process*. Close to the critical point, small changes in pressure or temperature result in large changes in density, allowing many properties to be tuned. Carbon dioxide and water are the most commonly used supercritical fluids (SCFs). For instance, above the critical point [304.1 K and 7.38 MPa (73.8 bar)], carbon dioxide is in the region of supercritical fluid, which is clearly demonstrated by a phase diagram (Figure 7.5). SCFs exhibit characteristics of both a liquid and a gas. In SCFs, the interface between solids and fluids lacks surface tension, which confers high viscosities on SCFs. Therefore, chemical compounds that have low solubility under ambient conditions can be easily dissolved in SCFs. Recently, adapting of high-temperature SCFs for producing nanoscale materials was demonstrated in near-critical or supercritical water (SCW). Arai et al. [42–44] used metal nitrates as precursors to prepare a wide variety of metal oxides in SCW. However, in their cases, the as-prepared particles were in agglomeration and coalescence. Ziegler et al. [45] has overcome the problem by adding alkanethiol ligands to the SCW reaction chamber. SCFs offer several advantages over conventional ways such as low surface tension, low viscosity, and high diffusivity.

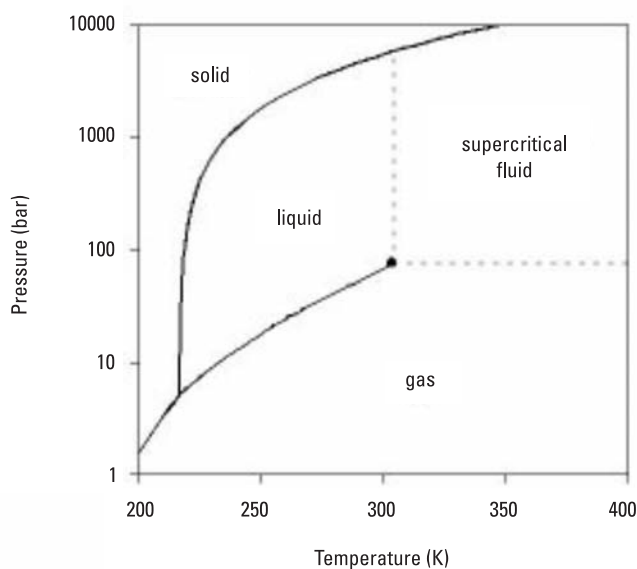


Figure 7.5 The pressure–temperature phase diagram of CO₂.

7.1.5 Templated Synthesis

Mesoporous materials and anodized aluminum oxide (AAO) with uniform pore sizes can be good templates for synthesizing nanoparticles. These templates can also be considered to be *nanoreactors*, in which chemical reactions take place. Their confined pore sizes will guide the NPs and can deliver a controllable size distribution to the final product. For instance, the mesoporous materials can shape the NPs in sizes ranging from 2 to 50 nm. Generally, the introduction of semiconductors into mesoporous materials can produce regular NPs with well-controlled particle sizes and distributions.

Two methods are used to load semiconductor NPs into pores of a mesoporous material: in situ or post-treatment. The former method is accomplished by mixing the NPs precursors with the micelles before the formation of the mesopores; the latter by directly grafting the NPs onto the pore surfaces of an as-prepared mesoporous material. Multiple means of post-treatment have been successfully explored, including surface grafting via sorption, phase transitions, ion exchange, complexation, and covalent grafting. Wang et al. [46] reported incorporation of CdS NPs inside SBA-15 mesoporous silica via ion exchange. The average particle size calculated from XRD data was about 3 nm, along with a blue-shift in UV-visible adsorption spectra. CdS NPs could also be easily deposited into mesoporous TiO₂ by chemical bathing [47]. Moreover, CdSe NPs bounded to the inside tunnels of mesoporous films [48] or spheres [49] share a uniform morphology with a narrow size distribution. It is also feasible to synthesize inorganic semiconductor NPs via phase transfer and phase separation approaches. For instance, Wang et al. [50] claimed that using facile liquid–solid–solution (LSS) could prepare monodispersed semiconductor NPs covering a wide bandgap, such as TiO₂, CuO, ZrO₂, SnO₂, CdS, Ag₂S, ZnS, PbS, MnS, ZnSe, and CdSe.

7.1.6 NPs of Organic Semiconductors

Compared to inorganic semiconductors, organic ones are fairly easy to synthesize and mechanically flexible. Organic semiconducting NPs of monomers (single molecules), oligomers (monomers linked to form a short chain), and polymers (monomers linked with a long chain) are all revealed. Examples of these NPs include acenes, oligothiophenes, polythiophenes, and poly(*p*-phenylene vinylene), as well as polyacetylene and their derivatives. Major mechanisms of semiconducting include conducting through Π -electrons or through unpaired electrons. It has been demonstrated that not only can a wide variety of physical properties provided by functional compounds be fully exploited, but these properties can also be modulated by the particle size and shape.

However, predicting how their properties will be affected by their complex, nanostructured morphology is still complicated. Generally, the electronic and

optical properties of organic NPs are fundamentally different from those of inorganic NPs, due to the weak van der Waals interaction between molecules. Uncovering and mapping the relationship between these properties and their structure require special attention to be paid when preparing monodispersed organic NPs.

The most convenient way to prepare organic NPs is the *reprecipitation method* [51–54]. In this method, a dilute solution of starting material in a water-soluble media is injected into vigorously stirred water. A sudden solubility change in the starting material causes the solutes to be precipitated in the form of nanocrystals. Xiao et al. [51] synthesized NPs of 1,3-diphenyl-5-(2-anthryl)-2-pyrazoline (DAP) by rapidly injecting the DAP from an acetonitrile stock solution to water. Mixing of acetonitrile with the water rapidly changes the dispersing properties of the solute and induces nucleation and growth of DAP NPs. Average particle size can be controlled through a ripening process via the regulation of the aging time. As-prepared DAP NPs showed a size-dependent property. This is probably caused by the electronic coupling between the confined molecules in the NPs.

The heterogeneous environment in which this precipitation method takes place makes the nucleation in the initial stage and the subsequent fast growth hard to control. This leads to a broad distribution in both the size and the shape of NPs. In contrast, a homogeneous solution makes size control much easier. For instance, Kang et al. [55] employed a colloid method to synthesize perylene NPs based on the reduction of perylene perchlorate by bromine anions in acetonitrile. NP growth was controlled by changing the monomer concentration or using a mixed solvent. Synthesized NPs can have sizes ranging from 25 to 95 nm with a dispersity of less than 10% (Figure 7.6).

Porphyrin and its derivatives also belong to the class of organic semiconductors, with remarkable catalytic, photo-, electro-, and biochemical properties. They are extensively used in self-assembly to prepare monolayers and thin films.

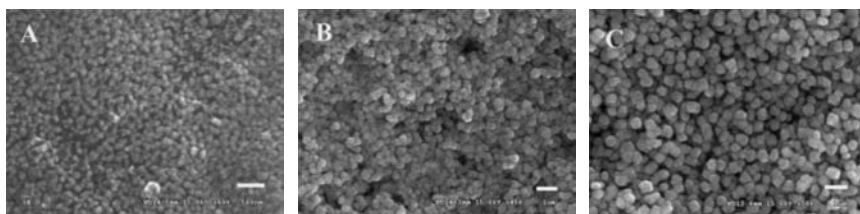


Figure 7.6 Typical FESEM images of quasispherical perylene NPs with a size of (a) 25 nm at $N = 0.3$, (b) 60 nm at $N = 0.6$ fabricated by the droplet injection of CTAB, and (c) 90 nm at about $N = 1.0$ fabricated by the single injection of CTAB, where $N = N(\text{CTAB})/N(\text{perylen perchlorate})$. All scale bars are 200 nm [55]. (Reprinted with permission from [55]. Copyright 2007 American Chemical Society.)

Gong et al. [56] reported the synthesis of porphyrin NPs with enhanced stability and catalytic rate, which contributed to the aggregate structure and the larger surface area. The average size of as-prepared porphyrin NPs is 27 nm, and the size distribution is remarkably narrow.

Pentacene is a small organic molecule consisting of five linearly condensed phenyl rings. Its extended conjugation, together with a favorable crystal structure confers it high hole mobility in the organic field effect transistor (OFET), comparable to amorphous silicon. Kita et al. [57] utilized the laser irradiation route to prepare pentacene NPs. They exposed pentacene in ethanol to a 532-nm Nd³⁺: YAG laser, and monitored the adsorption spectra change induced by laser irradiation. The resultant pentacene NPs were 4 to 13 nm in height and 10 to 70 nm in width.

Overall, we used this section to briefly introduce the synthetic strategies of preparing monodispersed inorganic and organic semiconductor NPs. The as-prepared NPs are widely used as building blocks to self-assemble into highly ordered micro- and nanostructures. Surface functionalization of these NPs surfaces, with suitable mono- or multifunctionalities paves the ways toward good self-organization [58].

7.2 Self-Assembly of Nanoparticles

Precise assembly and patterning of NPs are essential for creating functional devices for nanoelectronics and sensors. The noncovalent interactions that dictate self-assembly include hydrogen bonding, dipole–dipole, electrostatic, van der Waals, and hydrophobic interactions. In the following section, several bottom-up methods for self-assembling NPs are introduced.

7.2.1 Hydrogen Bonding-Based Assembly

Of the noncovalent interactions just listed, electrostatic and hydrogen bonding interactions are the mostly used approaches. To aid with these approaches, the surfaces of NPs are usually coated with a layer of surfactants or polymers. These coatings not only provide steric hindrance between neighboring NPs, but also render the NPs able to form hydrogen bonds between the terminal groups of the surface coatings. Thus, the NPs can be considered to be “bricks,” whereas the ligands or polymers are the “mortar” [59].

One of the typical examples of surfactants is dendrimer. Due to its huge unwrapped branches, dendrimer offers sufficient steric hindrance with easily tunable terminal groups. Through the dendrimer-mediated self-assembly, Boal et al. [60] reported a novel way to assemble γ -Fe₂O₃ NPs. First, γ -Fe₂O₃ NPs of 6.5 nm diameter were synthesized, and then they were anchored with thymine-

functionalized diol ligand by a place-exchange reaction. Upon assembly, a decrease in blocking temperature was observed, which was attributed to the increased spacing between the particles. Thus, the controlled manipulation of interparticle spacing provided new opportunities to vary the optical and magnetic properties of functional materials.

The interactions between antibody and antigen or between streptavidin and biotin are also good candidates for NP organization. Using hydrogen bonding between complementary single-stranded DNA is one such example. The rigidity and stability of DNA provides robust linkage atop NP surfaces. Chemical functionalization and precise processing of DNA materials, with very specific enzymes such as ligases and endonucleases, give rise to accurate tailoring of the DNA-based nanostructures.

7.2.2 Electrostatic Assembly

Apart from the hydrogen bonding and the van der Waals force, electrostatic interactions offer a simple and flexible pathway for creating well-organized, layer-by-layer (LbL) assembly of nanostructures. LbL assembly, largely developed by Decher et al. in the 1990s [61], is a popular and versatile technique for the fabrication of multilayered organic–inorganic films based on the alternating adsorption of oppositely charged materials. Potential applications of these materials have been exploited in areas such as surface modification, drug delivery, electrochemical devices, fuel cells, chemical sensors, nanomechanical sensors, and nanoscale chemical and biological reactors [62]. As illustrated by Figure 7.7, LbL assembly, can take place based on the electrostatic interactions that occur. Other types of driving forces, such as hydrogen bonding or covalent bonding, can be also applied to LbL.

Many sensing devices usually require the assembled nanostructure to be freestanding. Freestanding structures have no need for full contact with a solid substrate to sustain their shapes. Freestanding LbL microcapsules, microtubules,

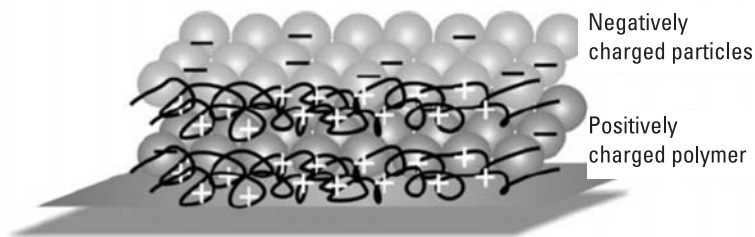


Figure 7.7 Electrostatic assembly of charged NPs on polymer using the LbL approach [63]. (Reprinted with permission from [63]. Copyright 2008 Wiley Interscience.)

microcubes, microcantilevers, and planar films have been prepared. The process usually starts with a “top-down” approach such as spin-coating, spraying, photolithography, electrochemical deposition, or microcontact printing to help the formation of freestanding LbL structures. For instance, flexible microcantilevers with a thickness of 170 nm have been fabricated from six alternating layers of clay platelets and $\text{Fe}_3\text{O}_4/\text{Fe}_2\text{O}_3$ NPs [64]. This approach produced arrays of microcantilevers on a substrate within a single processing cycle, which is crucial for mass microsensor manufacturing. Nanotubes containing TiO_2 and CdS quantum dots are also prepared by the LbL approach, which exhibits a novel emission band in the blue-wavelength range [65]. In another example, AAO was negatively charged by adsorbing poly(sodium 4-styrene sulfonate) (PSS), and then positively charged poly(diallyldimethylammonium chloride) (PDADMAC) was adsorbed onto the PSS surface as the second layer. The bilayer film (PSS/PDADMAC) provides a uniformly charged surface for the subsequent deposition of titanium precursor and CdS NPs. The reported strategy is useful for fabricating NP–NP composite nanostructure arrays. Monodispersed semiconducting perylene NPs can also be assembled into 3D structures, such as nanobelts or square nanorods [55]. The LbL-based self-organization process was identified from the TEM images in which a thin nanosheet is found to peel off from over a thick nanobelt. In some cases, the organic polyelectrolyte binders between each inorganic layer seem unnecessary, and they can be eliminated by calcinations [66].

7.2.3 Shape-Selective Assembly

Recently, the facet-selective interactions have been found interesting for NP organization. Via such an approach, most inorganic semiconductor NPs assemble into highly ordered structures. Nonspherical NPs can be easily prepared in different shapes, such as tetrahedral, octahedral, and cubic (see the above section). Because their shapes are so different, dipole–dipole interactions play an important role in the organization [(Figure 7.8(a)]. Because of the polarity differences between different facets of the particles, dipole moments are generated within the NPs [(Figure 7.8(b)]. In this kind of self-assembly, the energy difference between the crystallographic planes is the driving force.

It is of interest to know how NPs shape and dipole interactions combine to contribute to the self-assembly of dipolar NPs. In particular, because the dipole can orient in several directions relative to a cube face, the packing behavior observed differs from that of spheres. Zhang et al. [69] investigated the dipole-induced self-assembly of nanocubes with and without attractive face–face interactions by computer simulation. They concluded that nanocubes with $\langle 100 \rangle$ dipoles form straight wires, which agreed with the experiments; nanocubes with $\langle 110 \rangle$ dipoles are predicted to form sheets. The effect of face–face attractions

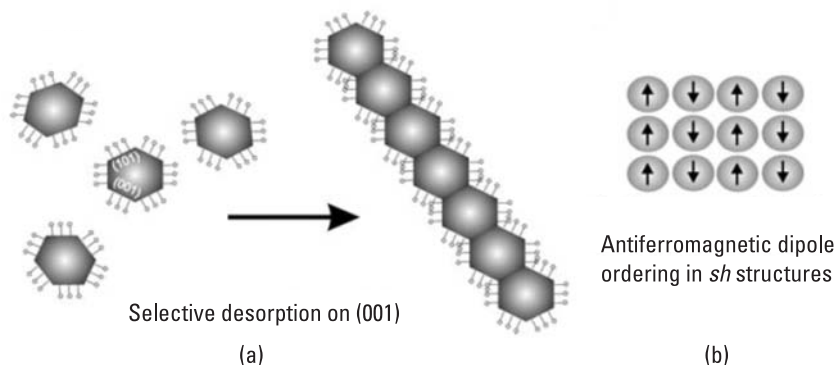


Figure 7.8 (a) Face-selective assembly [63, 67] and (b) dipole ordering in nanostructures [63, 68]. (Reprinted with permission from [63]. Copyright 2008 Wiley Interscience.)

was also examined, and they predicted that strong attractions would force nanocubes into stacked wires and sheetlike configurations, whereas a weak attraction would not alter the assembled wire structures.

In NP self-assembly via dipole–dipole interactions, ligands are always attached to the different facets by the affinity differences to generate anisotropy. The adsorption and desorption of ligands depends on the type of crystallographic facet involved. Using the anisotropy manifested in adsorption and desorption, many groups have systematically investigated the assembly of NPs in different dimensions. For instance, using oleic acid as a stabilizer, Yang et al. [70] reported shape and size controlled synthesis of ceria nanocubes bounded by six {200} planes through a rational one-pot approach, and assembled them on substrate in two and three dimensions. The strong dipole–dipole interactions between the polar {200} surfaces are probably the driving force for self-assembly of the nanocubes, because of the difficulty of controlling long-range ($d > 10$ nm) van der Waals forces. Cho et al. [71] organized PbSe NPs into nanowires and nanorings. The surfactant mixtures were utilized to mediate growth and achieve uniform nanowires. PbSe NPs could bind to each other via {100}, {110}, or {111} faces, depending on the surfactant molecules present in the reaction solution. Through different face connections, zigzag, helical, and straight nanowires can be synthesized.

Mixing NPs of different phases or sizes leads to more complex nanostructures. Phase separation of NPs occurs in the mixing of differently sized NPs, where bimodal packing of particles is observed at a certain mixing ratio. Chen et al. [72] mixed a series of NPs of CdTe and CdSe into highly ordered arrays. Moreover, by varying the molar ratio of the ligands, a self-assembled mixed monolayer of ligand undergoes a phase diagram-like evolution, forming ordered

ring structures composed of one or the other ligand [73]. In such highly curved molecular monolayer patterns, two polar singularities have subsequently been targeted and grafted with a single 11-mercaptoundecanoic acid (MUA) ligand, which is an elegant approach to forming linear chains of typically 10 to 20 isometric NPs by interparticle amidation with 1,6-diaminohexane.

7.2.4 Hydrophobic Assembly

The hydrophobic effect, the tendency for oil and water to segregate, is important in diverse phenomena, from the cleaning of laundry, to the creation of microemulsions to make new materials, to the assembly of proteins into functional complexes. This effect is multifaceted depending on whether hydrophobic molecules are individually hydrated or driven to assemble into larger structures. Despite the basic principles underlying the hydrophobic effect being qualitatively well understood, only recently have theoretical developments begun to explain and quantify many features of this ubiquitous phenomenon [74].

Hydrophobic interactions are often viewed as entropic, because they increase in strength with elevated temperature. Rycenga et al. [75] utilized such a driving force to assemble silver nanocubes into 1D, 2D, and 3D highly ordered superlattices, through selective functionalization of their face with hydrophobic ligands (Figure 7.9). They modified Ag NP surfaces with self-assembled monolayers (SAMs). The assembled structures are expected if the Ag nanocubes come into contact with their hydrophobic faces adjacent and parallel. In the assembly process, the hydrophobic driving force is proportional to the salvation free energy of the water surrounding the SAM-functionalized Ag nanocubes. The salvation free energy is primarily enthalpic and dependent on the number of hydrophobic and hydrophilic faces of each cube. The Ag nanocubes spontaneously assemble to reduce their salvation free energy by reducing the amount of hydrophobic surface area that is exposed to water. Interestingly, they demonstrated that the 3D superlattice can be assembled over length scales larger than 500 nm. This method can be extended to selectively modify the surface of as-prepared semiconductor NPs with hydrophobic molecules, with the spontaneous arrangement of NPs into highly ordered 1D, 2D, or 3D structures.

7.2.5 Template-Assisted Assembly

Inorganic and organic templates allow for structurally controlled assembly. Hard inorganic templates such as AAO nanoholes, inorganic nanowires, mesoporous materials, or carbon nanotubes provide greater toughness for assembling NPs and preserving their structural characteristics. The porous structures act as nanoreactors that can be applied to the creation of a variety of nanostructure arrays. By using these templates, large-area arrays of parallel wires can be obtained.

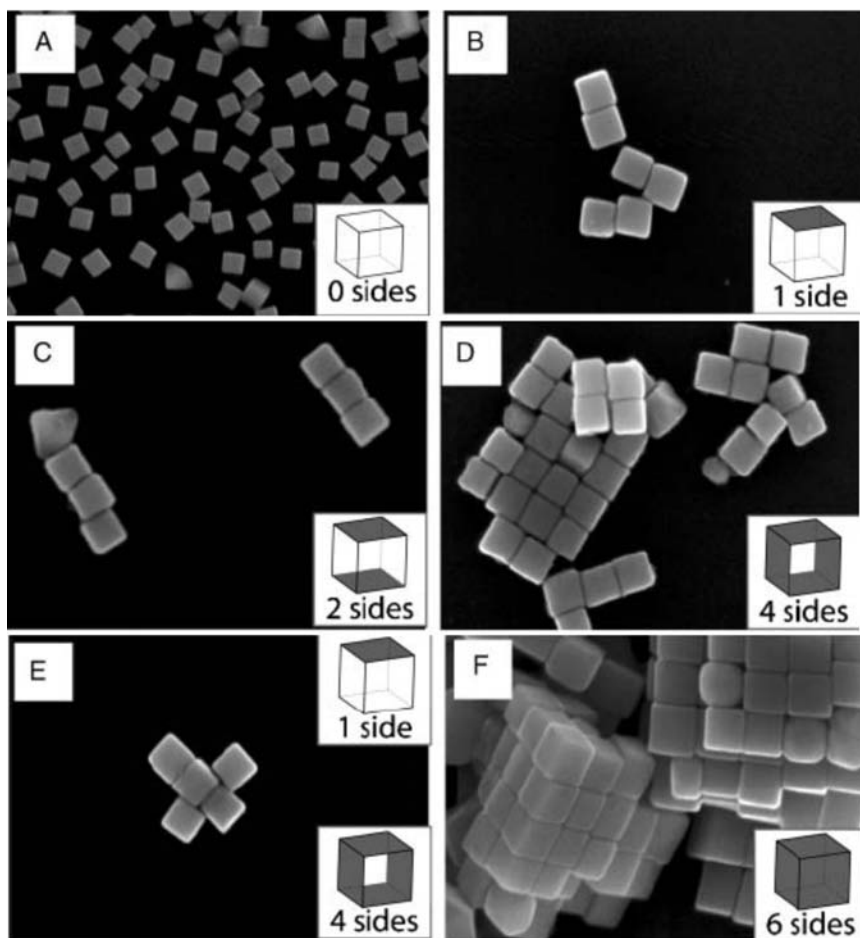


Figure 7.9 SEM images of Ag nanocubes and the assemblies. (a) As-prepared Ag NPs and (b–f) nanocubes whose faces have been selectively functionalized by hydrophilic and hydrophobic thiolate SAMs and then allowed to assemble in water. The darkened facets of the nanocubes rendered hydrophobic are shown in the bottom right corner of each panel, whereas the remaining faces are rendered hydrophilic [75]. (Reprinted with permission from [75]. Copyright 2008 Wiley Interscience.)

On the other hand, biological templates, such as DNA strands, peptides, and viruses are frequently investigated as the templates for assembling NPs. These templates provide a platform to organize particles through covalent and non-covalent interactions.

NPs can be assembled at the interior or exterior of the templates. Various examples are reported such as assembly of TiO_2 [76], CdSe [77], and CdTe [78] NPs. Recently, Yan et al. [79] reported the assembly of freestanding “particle

fiber,” using ice as a template. The NP was frozen in an alcohol bath exposed over a vapor of liquid nitrogen. When ice forms, all other ingredients previously dispersed were expelled, leaving many gaps either inside or between the ice nuclei. Because these gaps also grow in a predetermined unidirectional fashion, freestanding forests of 1D string-like features form after all of the NPs fall into these gaps (Figure 7.10).

7.2.6 Collective Properties of Self-Assembled Nanoparticles

The collective properties of self-assembled NPs are different from isolated particles and bulk phase. These properties may depend on the shape and the NP assembly at the mesoscopic scale. Readers can consult [80] for further details. When NPs are organized in 2D superlattices, collective optical and magnetic properties can be observed as a result of the dipolar interactions. For example, the optical properties of self-organized NPs give rise to several plasmon resonance modes, which are attributed to the anisotropy. In terms of magnetic properties, when NPs are deposited in compact hexagonal networks, the hysteresis loop is squarer than that of isolated NPs. The hysteresis loops for a chain-like structure are squarer than that of a well-ordered array, because the linear chains of NPs

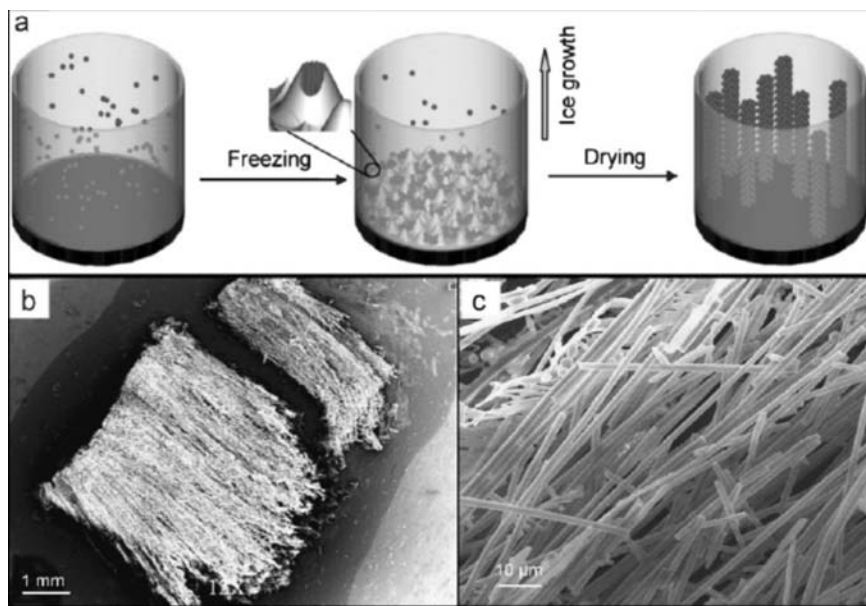


Figure 7.10 (a) Scheme of the ice freezing/drying process to prepare 1D fibers; SEM images of (b) a bundle of these structures with a length of 4.0 nm and (c) a zoomed image of unidirectionally grown fiber [79]. (Reprinted with permission from [79]. Copyright 2009 Wiley Interscience.)

behave as homogeneous nanowires. The ordering of NPs at the mesoscopic scale, in 3D superlattices, is not simply an aesthetic arrangement, but is, in fact, a new generation of materials.

7.3 Conclusion

In this chapter, we discussed strategies for synthesizing both inorganic and organic semiconductor NPs, especially those methods related to preparing mono-dispersed NPs. The as-prepared semiconductor NPs can be used as building blocks to spontaneously organize into highly ordered structures in one, two, or three dimensions, driven by the weak noncovalent interactions, such as hydrogen bonding, dipole–dipole, electrostatic, van der Waals, and hydrophobic interactions. These self-assembled NPs display collective properties different from the isolated particles and bulk phases. We merely demonstrated limited strategies and phenomena about the synthesis and self-assembly, but a large number of work remains beyond our line of sight.

References

- [1] Cushing, B. L., V. L. Kolesnichenko, and C. J. O'Connor, *Chemical Reviews*, Vol. 104, 2004, p. 3893.
- [2] Wang, Y., et al., *Inorganic Chemical Communications*, Vol. 5, 2002, p. 751.
- [3] Borse P. H., et al., *J. of Materials Science: Materials in Electronics*, Vol. 13, 2002, p. 553.
- [4] Niederberger, M., and G. Garnweitner, *Chemistry—A European J.*, Vol. 12, 2006, p. 7282.
- [5] Sun, S., and H. Zeng, *J. of the American Chemical Society*, Vol. 124, 2002, p. 8204.
- [6] Rockenberger, J., E. C. Scher, and A. P. Alivisatos, *J. of the American Chemical Society*, Vol. 121, 1999, p. 11595.
- [7] Kumar, S., and T. Nann, *Small*, Vol. 2, 2006, p. 316.
- [8] Murray, C. B., D. J. Norris, and M. G. Bawendi, *J. of the American Chemical Society*, Vol. 115, 1993, p. 8706.
- [9] Manna, L., E. C. Scher, and A. P. Alivisatos, *J. of the American Chemical Society*, Vol. 122, 2000, p. 12700.
- [10] Wang, Z., et al., *Advanced Functional Materials*, Vol. 16, 2006, p. 661.
- [11] Jun, Y. W., et al., *J. of the American Chemical Society*, Vol. 123, 2001, p. 5150.
- [12] Cozzolli, P. D., et al., *Chemistry of Materials*, Vol. 17, 2005, p. 1296.
- [13] Hench, L. L., and J. K. West, *Chemical Reviews*, Vol. 90, 1990, p. 33.
- [14] Brinker, C. J., and G. W. Scherer, *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, London: Academic Press, 1990.

-
- [15] Niederberger, M., *Accounts of Chemical Research*, Vol. 40, 2007, p. 793.
- [16] Zhang Z., et al., *Angewandte Chemie International Edition*, Vol. 44, 2005, p. 3466.
- [17] Li, X. L., et al., *Chemistry—A European J.*, Vol. 12, 2006, p. 2383.
- [18] Li, Z., Q. Sun, and M. Y. Gao, *Angewandte Chemie International Edition*, Vol. 44, 2005, p. 123.
- [19] Yang, T., et al., *J. of Physical Chemistry B*, Vol. 109, 2005, p. 232333.
- [20] Zitoun, D., et al., *J. of the American Chemical Society*, Vol. 127, 2005, p. 15034.
- [21] Zhong, X., et al., *J. of Physical Chemistry B*, Vol. 110, 2006, p. 2.
- [22] Yu, T., et al., *Angewandte Chemie International Edition*, Vol. 44, 2005, p. 7411.
- [23] He, T., et al., *Chemistry of Materials*, Vol. 17, 2005, p. 4023.
- [24] Ghosh, M., et al., *J. of Materials Chemistry*, Vol. 16, 2006, p. 106.
- [25] Park, J., et al., *Advanced Materials*, Vol. 17, 2005, p. 429.
- [26] Joo, J., et al., *Advanced Materials*, Vol. 17, 2005, p. 1873.
- [27] Yuhas, B. D., et al., *Angewandte Chemie International Edition*, Vol. 45, 2006, p. 420.
- [28] Liu, Q., et al., *J. of the American Chemical Society*, Vol. 127, 2005, p. 5276.
- [29] Epifani, M., et al., *Chemistry of Materials*, Vol. 17, 2005, p. 6468.
- [30] Woo, K., et al., *Inorganic Chemistry*, Vol. 44, 2005, p. 7171.
- [31] Yin, H., H. P. Too, and G. M. Chow, *Biomaterials*, Vol. 26, 2005, p. 5818.
- [32] Kirchner, C., et al., *Nano Letters*, Vol. 5, 2005, p. 331.
- [33] Trentler, T. J., et al., *J. of the American Chemical Society*, Vol. 121, 1999, p. 1613.
- [34] Joo, J., et al., *J. of the American Chemical Society*, Vol. 125, 2003, p. 6553.
- [35] Fanelli, A. J., and J. V. Burlew, *J. of the American Ceramic Society*, Vol. 69, 1986, p. C174.
- [36] Pinna, N., et al., *Advanced Materials*, Vol. 16, 2004, p. 2196.
- [37] Ivanda, M., et al., *J. of Molecular Structure*, Vol. 481, 1999, p. 465.
- [38] Hoar, T. P., and J. H. Schulman, *Nature*, Vol. 152, 1943, p. 102.
- [39] Ingert, D., and M. P. Pileni, *Advanced Functional Materials*, Vol. 11, 2001, p. 136.
- [40] Pinna, N., et al., *Langmuir*, Vol. 17, 2001, p. 7982.
- [41] Simmons, B. A., et al., *Nano Letters*, Vol. 2, 2002, p. 263.
- [42] Adschiri, T., K. Kanazawa, and K. Arai, *J. of the American Ceramic Society*, Vol. 75, 1992, p. 1019.
- [43] Adschiri, T., Y. Hakuta, and K. Arai, *Industrial & Engineering Chemistry Research*, Vol. 29, 2000, p. 4091.

- [44] Hakuta, Y., et al., *J. of the American Ceramic Society*, Vol. 81, 1998, p. 2461.
- [45] Ziegler, K. J., et al., *J. of the American Chemical Society*, Vol. 123, 2001, p. 7797.
- [46] Wang, S., D. G. Choi, and S. M. Yang, *Advanced Materials*, Vol. 16, 2002, p. 1311.
- [47] Piris, J., et al., *J. of Physical Chemistry C*, Vol. 112, 2008, p. 7742.
- [48] Wark, M., H. Wellmann, and J. Rathousky, *Thin Solid Films*, Vol. 458, 2004, p. 20.
- [49] Wang, P., et al., *Acta Materialia*, Vol. 56, 2008, p. 1144.
- [50] Wang, X., et al., *Nature*, Vol. 437, 2005, p. 121.
- [51] Xiao, D., et al., *J. of the American Chemical Society*, Vol. 125, 2003, p. 6740.
- [52] Horn, D., and J. Rieger, *Angewandte Chemie International Edition*, Vol. 40, 2001, p. 4331.
- [53] Hu, J. S., et al., *J. of the American Chemical Society*, Vol. 127, 2005, p. 17090.
- [54] Xiao, D. B., et al., *J. of the American Chemical Society*, Vol. 126, 2004, p. 15439.
- [55] Kang, L., et al., *J. of the American Chemical Society*, Vol. 129, 2007, p. 7305.
- [56] Gong, X., et al., *J. of the American Chemical Society*, Vol. 124, 2002, p. 14290.
- [57] Kita, S., et al., *Japan J. of Applied Physics*, Vol. 45, 2006, p. 6501.
- [58] Baraton, M. I., *Synthesis Functionalization and Surface Treatment of Nanoparticles*, Stevenson Ranch, CA: American Scientific Publishers, 2003.
- [59] Arumugam, P., et al., *Polymer International*, Vol. 56, 2007, p. 461.
- [60] Boal, A. K., et al., *Chemistry of Materials*, Vol. 16, 2004, p. 3252.
- [61] Decher, G., *Science*, Vol. 277, 1997, p. 1232.
- [62] Jiang, C., and V. V. Tsukruk, *Advanced Materials*, Vol. 18, 2006, p. 829.
- [63] Kinge, S., M. C. Calama, and D. N. Reinhoudt, *Chem. Phys. Chem.*, Vol. 9, 2008, p. 20.
- [64] Hua, F., T. Cui, and Y. M. Lvov, *Nano Letters*, Vol. 4, 2004, p. 823.
- [65] Guo, Y. G., J et al., *Advanced Functional Materials*, Vol. 15, 2005, p. 196.
- [66] Lee, D., M. F. Rubner, and R. E. Cohen, *Nano Letters*, Vol. 6, 2006, p. 2305.
- [67] Polleux, J., et al., *Chemistry—A European J.*, Vol. 11, 2005, p. 3541.
- [68] Talapin, D. V., et al., *Nano Letters*, Vol. 7, 2007, p. 1213.
- [69] Zhang, X., Z. Zhang, and S. C. Glotzer, *J. of Physical Chemistry C*, Vol. 111, 2007, p. 4132.
- [70] Yang, S., and L. Gao, *J. of the American Chemical Society*, Vol. 128, 2006, p. 9330.
- [71] Cho, K. S., et al., *J. of the American Chemical Society*, Vol. 127, 2005, p. 7140.
- [72] Chen, Z., and S. O'Brien, *Nano*, Vol. 2, 2008, p. 1219.

- [73] Vries, G. A., et al., *Science*, Vol. 315, 2007, p. 358.
- [74] Chandler, D., *Nature*, Vol. 437, 2005, p. 640.
- [75] Rycenga, M., J. M. McLellan, and Y. Xia, *Advanced Materials*, Vol. 20, 2008, p. 2416.
- [76] Li, X., et al., *J. of Physical Chemistry B*, Vol. 107, 2003, p. 2453.
- [77] Haremsza, J. H., et al., *Nano Letters*, Vol. 2, 2002, p. 1253.
- [78] Banerjee, S., and S. S. Wong, *Advanced Materials*, Vol. 16, 2004, p. 34.
- [79] Yan, J., et al., *Advanced Materials*, Vol. 21, 2009, p. 314.
- [80] Pileni, M. P., *Accounts of Chemical Research*, Vol. 40, 2007, p. 685.

Part III

Nanomaterials Properties and Applications

In this last part of the book, we are going to explore the unique properties and applications of nanomaterials. We will focus our topics on nanoelectronic, nanostructural, and biomaterials because these are the frontier of nanomaterials science and engineering.

8

Nanoelectronic Materials

The continuous scaling of device size in complementary metal-oxide semiconductor (CMOS) transistor devices has been the fundamental driving force of semiconductor and microelectronics industry in last several decades. Taking challenges from Moore's law [1], scientists and engineers have been able to relentlessly push down the limits—today, we are able to pack more than 1.9 billion transistors into one chip compared with 60 transistors in the same area 40 years ago.

However, even Gordon Moore himself acknowledged that his law cannot be sustained forever. He noted in 2005 [2] that eventually transistor size will reach the limit of atomic size, which is the fundamental barrier to scaling. Like most people, he also believed that we would have one or two more generations before traditional CMOS devices can no longer be scaled down, at which point, a new device will be introduced and microelectronics will be truly converted to nanoelectronics. Now we are quickly approaching this limit.

In this chapter, we introduce several areas that have attracted much attention in recent years including single-electron transistors, carbon nanotube-based devices, and spintronics.

8.1 Single-Electron Transistors (SETs)

A single-electron transistor is a device that can control the motion of a single electron with quantum dots and a tunnel junction. It is a new type of device with no scalability issues because it is based on nanomaterials science and the quantum effect. In this section, we first introduce the fundamental principle of a single-electron capacitor and transistor based on the Coulomb blockade

and quantum well confinement and then discuss the practical techniques used to fabricate SETs.

8.1.1 Single-Electron Capacitor

Just like metal-oxide semiconductor field-effect transistors (MOSFETs) and MOS capacitors, to understand the fundamental principles of SETs, we need to start with a single-electron capacitor (SEC), which is the simplest known single-electron device. Also called a single-electron box, the capacitor usually consists of a quantum dot that is asymmetrically located between two electrodes as shown in Figure 8.1. The capacitance between the quantum dot and the closer source electrode is the tunnel junction, and the other electrode is called the control gate capacitor. Once we apply voltage across two electrodes, electrons will be injected into quantum dots or, vice versa, ejected from quantum dots through the tunneling junction depending on the signs of the voltages. This is the same for any charge storage media between two electrodes. The difference in the case of the single-electron box is that because of the size of quantum dots, every time we try to inject an additional electron into quantum dots, excessive energy is needed due to Coulomb blockade effect. This unique property enables us to control the motion of a single electron through tunnel junctions.

The Coulomb blockade is caused by an increase in the excessive charging energy due to the smaller size of quantum dots. Strictly speaking, it is not a quantum effect but rather a classic phenomenon under the effect of nanometer size.

For a sphere-shaped object, the relationship between the voltage and the charge can be shown as follows:

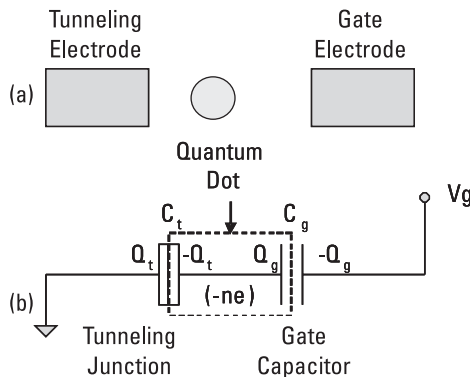


Figure 8.1 (a) Single-electron box and (b) its equivalent circuit.

$$V = \frac{Q}{4\pi\epsilon_0 r} \Rightarrow C_{sp} = 4\pi\epsilon_0 r \quad (8.1)$$

where r is the radius of the sphere and C_{sp} is the capacitance of this sphere to the infinite large plate. Therefore, the excessive energy required to add one electron to a sphere-shaped object is

$$E_c = \frac{1}{2} QV = \frac{1}{2} e \frac{e}{4\pi\epsilon_0 r} = \frac{1}{2} \frac{e^2}{4\pi\epsilon_0 r} = \frac{e^2}{2C_{sp}} \quad (8.2)$$

In a macroscopic world, because electron charge is a small number, this excessive energy is usually negligible. For example, for a sphere with a radius of 1m, the excessive energy for adding one extra electron is only 3.78E-10 eV, a really small number. However, once the dot size goes down to the nanometer region, for example, the excessive energy for a 10-nm quantum dot is 0.0378 eV, and for a 3-nm quantum dot it is 0.126 eV. Compared with thermal energy at room temperature kT , which is 0.025 eV, these excessive charging energies can no longer be ignored and overcome easily by thermal fluctuation. Therefore, addition energy is needed every time we want to add an extra electron into small quantum dots. Figure 8.2 shows excessive energy correlations with radius compared with kT . Clearly, at the nanometer region, this energy has to be considered.

As shown in the equivalent circuit in Figure 8.1(b), when the control gate voltage is zero, there is no electron tunnel through the junction. So the net charge in the quantum dot is zero. As gate voltage increases to a certain value, electrons on the source electrode are attracted toward the quantum dot and start to tunnel through the junction. However, once the first electron tunnels through, the net charge in the quantum dot becomes $1e$ and a further increase of electrons in the quantum dots will not happen immediately until the gate voltage further increases a certain value to overcome the excessive energy due to the Coulomb blockade. Let's calculate the relationship between gate voltage and net charge in quantum dots. Assume that we have a net charge of n electrons or $-ne$ in the quantum dots, and the gate voltage applied is V_g ; further assume that the charge and the capacitance between the source and the quantum dots are Q_t , C_t , and between the quantum dots and the gate are Q_g , C_g . The boundary condition for a single-electron box is then

$$\begin{aligned} Q_t - Q_g &= -ne \\ \frac{Q_t}{C_t} + \frac{Q_g}{C_g} &= V_g \end{aligned} \quad (8.3)$$

From this, we can easily derive the charge amount in Q_t and Q_g as

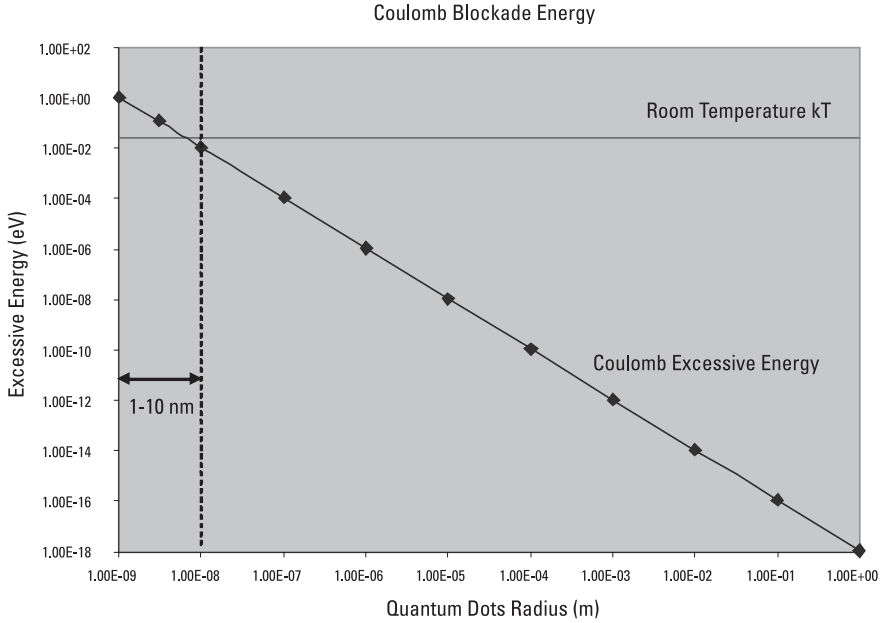


Figure 8.2 Excessive charging energy increases rapidly with the decrease of particle size—at the nanometer region, it can no longer be ignored by thermal fluctuations at room temperature.

$$Q_g = \frac{V_g C_t C_g + C_g n e}{C_g + C_t} \quad (8.4)$$

$$Q_t = \frac{V_g C_t C_g - C_t n e}{C_g + C_t}$$

Now if we consider the entire box as a system, the external work that has been done to the system can be calculated as W :

$$W = V_g Q_g = \frac{V_g^2 C_t C_g + V_g C_g n e}{C_g + C_t} \quad (8.5)$$

The excessive charging energy that need to overcome is E_c as described previously:

$$E_c = \frac{1}{2} \frac{Q_t^2}{C_t} + \frac{1}{2} \frac{Q_g^2}{C_g} = \frac{Q_t}{2} \frac{Q_t}{C_t} + \frac{Q_g}{2} \frac{Q_g}{C_g}$$

Substituting (8.3) into this equation, the function can be further reduced to

$$E_c = \frac{Q_t}{2} \left(V_g - \frac{Q_g}{C_g} \right) + \frac{Q_g}{2} \frac{Q_g}{C_g} = \frac{Q_t V_g}{2} + \frac{n e Q_g}{2 C_g}$$

Replacing results from (8.4) into the equation above, we can obtain the excessive charging energy:

$$E_c = \frac{V_g^2 C_g C_t + n^2 e^2}{2(C_g + C_t)} \quad (8.6)$$

Therefore, the system internal energy when the quantum dot has n electrons or a net charge of ne is the difference between external work done and the excessive charging energy that needs to be overcome:

$$E(n) = W - E_c = \frac{V_g^2 C_g C_t + 2neV_g C_g - n^2 e^2}{2(C_g + C_t)} \quad (8.7)$$

At this point, the system is at equilibrium with respect to the gate voltage V_g . To maintain the electron number of n , we will have to satisfy the minimum energy role as follows:

$$E(n) < E(n \pm 1) \quad (8.8)$$

To satisfy this condition, we obtain

$$\begin{aligned} E(n) &= \frac{e^2 n^2 - 2enC_g V_g - C_t C_g V_g^2}{2(C_t + C_g)} < E(n+1) \\ &= \frac{e^2 (n+1)^2 - 2e(n+1)C_g V_g - C_t C_g V_g^2}{2(C_t + C_g)} \\ e^2 n^2 - 2enC_g V_g &< e^2 (n+1)^2 - 2e(n+1)C_g V_g \\ 2eC_g V_g &< e^2 (2n+1) \Rightarrow V_g < \left(n + \frac{1}{2} \right) \frac{e}{C_g} \end{aligned} \quad (8.9)$$

When we consider both sides, the relationship between gate voltage V_g and the net charge is

$$\left(n - \frac{1}{2} \right) \frac{e}{C_g} < V_g < \left(n + \frac{1}{2} \right) \frac{e}{C_g} \quad (8.10)$$

Therefore, if we plot this relationship, the net charge in the quantum dot is actually a step function (also called a Coulomb staircase) of gate voltage of V_g as shown in Figure 8.3.

8.1.2 Operating Principles for SETs

A single-electron transistor is a three-terminal device based on the same principles as a single-electron capacitor. As shown in Figure 8.4, a SET has a source, drain, and gate electrode with a quantum dot located in the middle of the source and drain. The source is usually grounded and voltages are applied on the gate (V_g) and drain electrode (V_d). Figure 8.4(b) shows the schematic circuit of a SET. In this case, we have two tunnel junctions, one on the source and one on the drain side. These two tunnel junctions share the same quantum dot as the common electrode (also called an *island*). The potential of the quantum dot can be adjusted by the gate voltage V_g through the gate capacitance. When the gate voltage is zero, the quantum dot is in the equilibrium state with no electron tunnel through on both the source and drain sides. However, when the gate voltage exceeds the threshold voltage V_t , electrons will tunnel through from the source side into the quantum dot and then continuously eject into the drain electrode and thus form a current. This is why we call it a single-electron transistor.

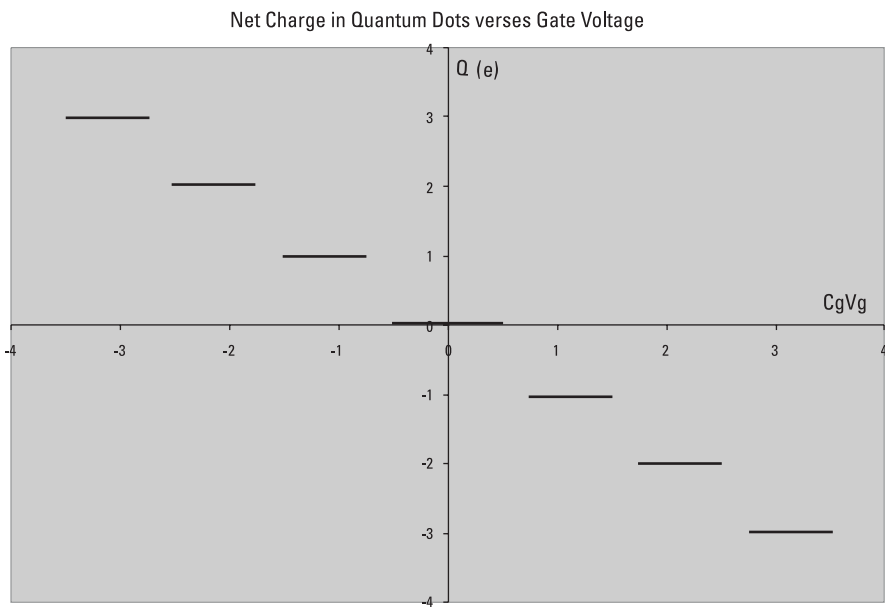


Figure 8.3 The net charge in quantum dots as a function of gate voltage—a clear indication that the Coulomb blockade is in effect. This is called a Coulomb staircase.

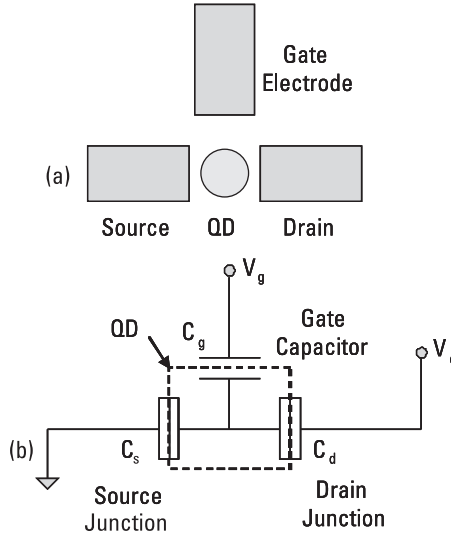


Figure 8.4 (a) Single-electron transistor and (b) its equivalent circuit.

To simplify the problem, we can consider this SET to be two independent tunnel junctions, each of which will have to satisfy the Coulomb blockade condition we derived in the previous section. If we look only at the source side tunnel junction [the equivalent schematic circuit is shown in Figure 8.5(a)], the source electrode is still grounded as in the single-electron box case. The gate and drain in the transistor case, however, effectively become the gate in the box case. The total charge on the effective gate side is $Q_{\text{eff}} = C_g V_g + C_d V_d$. The total effective gate capacitance can be considered to be the gate and drain capacitance running in parallel and, thus, $C_{\text{eff}} = C_g + C_d$. Therefore, the effective gate voltage can be calculated as $V_{\text{eff}} = Q_{\text{eff}} / C_{\text{eff}} = (V_g C_g + V_d C_d) / (C_g + C_d)$. Substituting both effective gate capacitance and voltage into (8.10), we obtain the equilibrium condition for the source side tunnel junction as follows:

$$\left(n - \frac{1}{2}\right) \frac{e}{C_g + C_d} < \frac{C_g V_g + C_d V_d}{C_g + C_d} < \left(n + \frac{1}{2}\right) \frac{e}{C_g + C_d}$$

and it can be further reduced to

$$\frac{1}{C_d} \left(ne - \frac{e}{2} - C_g V_g \right) < V_d < \frac{1}{C_d} \left(ne + \frac{e}{2} - C_g V_g \right) \quad (8.11)$$

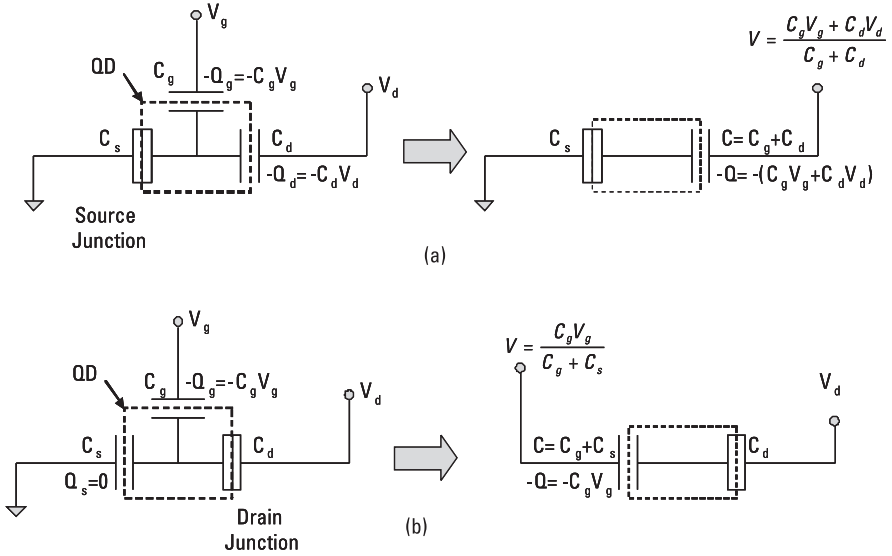


Figure 8.5 Illustration of equivalent circuit for (a) the source side and (b) the drain side tunnel junction in a SET device.

We can apply the same methodology on the drain side tunnel junction. The difference is that now the drain side is not grounded but the source is. So the total charge now is only $Q_{\text{eff}} = C_g V_g$. The effective capacitance is still $C_{\text{eff}} = C_g + C_s$. Now since the drain electrode is not grounded, the effective voltage applied on this tunnel junction is only $V_{\text{eff}} = V_g C_g / (C_g + C_d) - V_d$. Again substituting these relationships back into (8.10), we have the following for the drain side:

$$\left(n - \frac{1}{2}\right) \frac{e}{C_g + C_s} < \frac{C_g V_g}{C_g + C_s} - V_d < \left(n + \frac{1}{2}\right) \frac{e}{C_g + C_s}$$

This is equivalent to the following relationship:

$$\frac{C_g V_g - ne + \frac{e}{2}}{C_g + C_s} > V_d > \frac{C_g V_g - ne - \frac{e}{2}}{C_g + C_s} \quad (8.12)$$

Given the relationships in (8.11) and (8.12) on the Coulomb blockade for both the source and the drain side tunnel junctions, we can display those relationships in a $V_g - V_d$ plot. For example, Figure 8.6(a) plots the Coulomb blockade condition for the source side tunnel junction. The number labeled between two lines is n , meaning that when V_g and V_d satisfy the condition

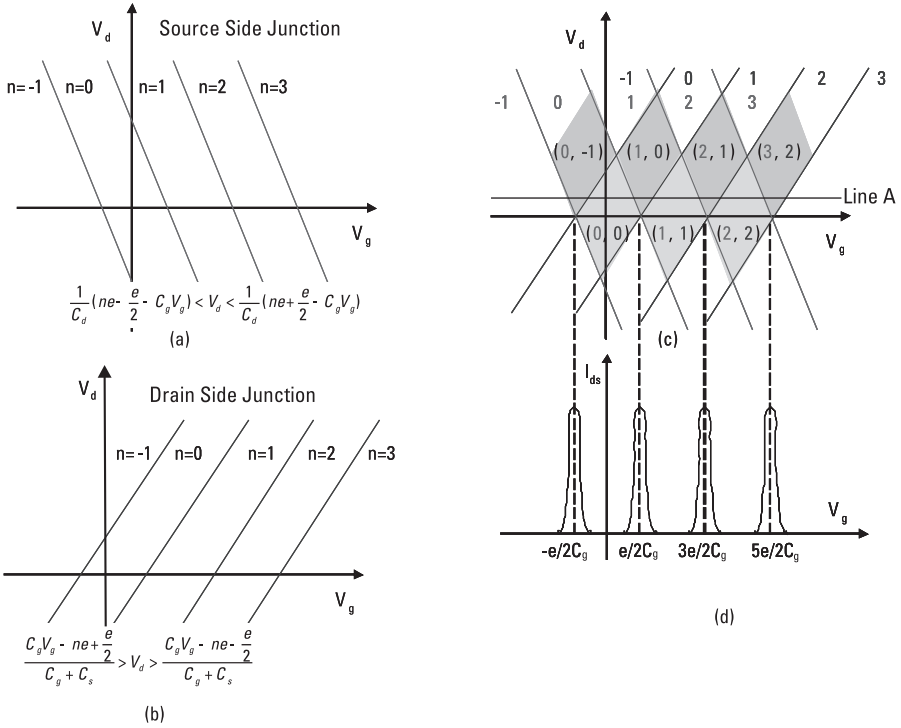


Figure 8.6 Coulomb blockade conditions for (a) source side and (b) drain side tunneling and fundamental operating principle for (c) SET and (d) Coulomb oscillator. (Adapted from [13].)

between the two lines, the source side tunnel junction requires n electrons inside the quantum dot. From this figure, we can easily see that if $V_g = V_d = 0$, no electron should be inside the quantum dot. If we keep $V_d = 0$ while increasing V_g , we will pass the region of $n = 1, n = 2, \dots$, meaning that the equilibrium number of electrons increases in the quantum dots. On the other hand, if we increase V_d while keep $V_g = 0$, electrons will tunnel from the quantum dots to the drain side and thus the net change becomes $-e$ (losing one electron). In summary, for any combination of (V_g, V_d) , Figure 8.6(a) shows the equilibrium number of electrons that will satisfy the Coulomb blockade condition from the source side tunnel junction.

The same plot can be made for the drain side tunnel junction as well [Figure 8.6(b)]. Again for a given (V_g, V_d) , the number shown here is the equilibrium number of electrons for the drain side tunnel junction. Because a SET can be considered to be two tunnel junctions that satisfy the Coulomb blockade simultaneously, both (8.11) and (8.12) will have to be satisfied. Therefore, Figure 8.6(c) is an overlay of Figures 8.6(a) and (b) together and shows

that an equilibrium condition can be obtained—this is the operating principle for single-electron transistors. As shown in Figure 8.6(c), the light gray area labeled with (0,0), (1,1), and (2,2) in the V_g versus V_d plot is where both the source and drain side tunnel junctions require the same equilibrium number of electrons. Therefore, for any given pairs of (V_g , V_d) inside these areas, the quantum dots will reach an equilibrium state with a fixed number of electrons as labeled in Figure 8.6(c). This is called the *Coulomb blockade region* where the transistor is off. On the other hand, in the dark gray area above, electrons can take at least two numbers, one for the source and one for the drain. For example, for a given pair of (V_g , V_d) in the region labeled (1,0), the source side tunnel junction required one electron for equilibrium so one electron will tunnel from the source to the quantum dot. However, the drain side tunnel junction required zero electrons inside the quantum dot so the electron that just tunneled through from the source side will continue to pass through the drain side tunnel junctions. The net result is that an electron tunnels from the source side through the quantum dot to the drain side. Therefore, a source-to-drain current can be observed—the transistor has been turned on.

Some special operating conditions of SETs can be discussed using Figure 8.6(c). For example, when $V_d = 0$, the transistor is essentially simplified to a capacitor. In this case, Figure 8.6(c) is the same as Figure 8.3 with each node as $e/2C_g$, $3e/2C_g$, $5e/2C_g$, with the increase of gate voltage, the number of electrons stored in quantum dots increased. However, there is no current flow from source to drain. The transistor will be completely in an “off” state due to the Coulomb blockade. On the other hand, when we apply a small V_d as labeled in line A, the current between the source and drain I_{ds} is still mostly zero until V_g closes the “node” locations ($e/2C_g$, $3e/2C_g$, $5e/2C_g$,). Only at those locations, where line A entered the dark gray region as described above and one electron tunnels through from the source to the drain, is I_{ds} observed and the transistor on. This oscillating I_{ds} versus V_g curve represents *Coulomb oscillation* [Figure 8.6(d)] and it is one of the key characteristics of a single-electron transistor.

On the other hand, when gate voltage $V_g = 0$, one cannot observe the source-to-drain current I_{ds} until V_d is increased to a certain value in the yellow region as shown in Figure 8.6(c). This is the “off” state for a SET. However, when $V_g = e/2C_g$, any V_d applied will lead to the source-to-drain current I_{ds} and thus the transistor is “on.” This is a fundamental operating principle for single-electron transistors.

8.1.3 Quantum Effect on SETs

Unlike a traditional field-effect transistor (FET), single-electron capacitors and transistors are built based on the principles of quantum mechanics. Electrons

tunnel from the source to the quantum dot (island) or from the quantum dot to the drain side. The tunneling itself is a quantum mechanical effect regardless of the Coulomb blockade. A typical tunneling diagram is shown in Figure 8.7(a). Two thin layers of dielectric sandwich an island (in this case, it is a quantum dot) between with one side being the source, the other the drain electrode. Because the dielectric layers are so thin, the electron at the Fermi level of the source can “tunnel” through the thin dielectric layers even though in the classic case it does not have sufficient energy to overcome the potential barriers. This is the typical tunneling based on quantum mechanics.

The Coulomb blockade, as discussed earlier, simply adds a set of discrete energy levels inside the quantum dot (island) due to the excessive charging energy as shown in Figure 8.7(b). The electron tunneling through from the source to the quantum dot cannot occupy any arbitrary energy state due to the Coulomb blockade. For example, as shown in Figure 8.7(b), all energy levels in the quantum dots that are lower than the Fermi level on the source side have been occupied. Therefore, the electrons in the source cannot tunnel through further unless we continuously apply voltage (raising the Fermi level on the source side). This is called the “block” state. As shown in Figure 8.7(b), all energy levels inside the quantum dots due to the Coulomb blockade are equally spaced just like in the case of the Coulomb staircase.

In addition to the Coulomb blockade, the quantum mechanics effect also plays a significant role with the decrease of quantum dot size up to a number comparable to the de Broglie wavelength of the electron. As discussed in Chapter 2, when electrons are injected into a quantum dot, which can be considered

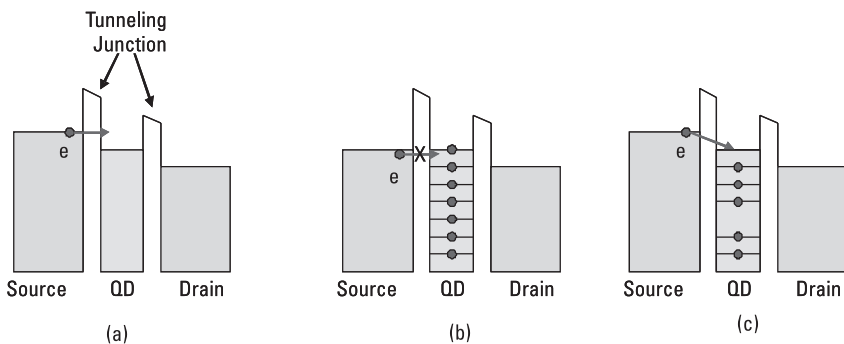


Figure 8.7 Illustration of (a) traditional tunneling in which electrons can occupy any energy level in the quantum dot once they tunnel through the dielectric junction; (b) situation in which electrons cannot tunnel through since they can only occupy a set of discrete energy levels in the quantum dot due to the Coulomb blockade, so it is in a “block” state; and (c) when we consider the quantum effect, the spacing between different energy levels in the quantum dot is no longer equal.

to be a 3D quantum well, they cannot take any arbitrary energy state because the energy levels inside the quantum dots are discrete or quantized. Furthermore, because electrons are fermions, only two electrons with different spins can share the same state. For example, if we use a 3D infinite quantum well as a rough estimation, the eigenvalue of energy follows (1.38) in Chapter 2:

$$E_{n_x, n_y, n_z} = \frac{h^2}{8mr^2} [n_x^2 + n_y^2 + n_z^2]$$

The ground state and the second state of energy levels are, according to (1.39) and (1.40),

$$E_{111} = \frac{3h^2}{8ma^2} \quad E_{211} = E_{121} = E_{112} = \frac{3h^2}{4ma^2}$$

Here the first energy level can accommodate two electrons, one spin up and one spin down, while the second energy level can contain up to six electrons due to degeneracy. Using 3-nm quantum dots as an example, the energy we discussed here is about 0.125 eV, which is comparable to the excessive energy from the Coulomb blockade.

Therefore, as shown in Figure 8.7(c), the energy levels inside the quantum dot really should be the sum of both the excessive charging energy due to the Coulomb blockade and the potential energy due to quantum confinement. As a result, the energy levels are raised and the spacing between each energy level is no longer the same. This will further require a higher gate voltage to turn on the transistor. Figure 8.8 shows Coulomb oscillation when we consider quantum effects. Compared with Figure 8.6(d), it is apparent that the gate voltage has been shifted. Because the electron ground energy is not zero in the infinite quantum well case, the gate voltage required for one electron to tunnel through is $e/2C_g + E_{111}/e$. The shift in gate voltage is $\Delta E/e$, where ΔE is E_{111} . The delta between the first and second node remains as e/C_g since the first electron and second electron share the same energy level of the ground state. However, for $n = 3$, because two electrons already occupy the lowest energy level at E_{111} , the third electron tunneling through not only has to overcome the excessive charging energy due to the Coulomb blockade but also the energy delta due to the quantum effect. Therefore, the gate voltage is shifted higher and the delta between the second and the third oscillations is $e/C_g + (E_{211} - E_{111})/e = e/C_g + \Delta E/e$. Because the second energy level is degenerate and can contain up to six electrons, the deltas between the third, fourth, . . . , eighth remain the same as e/C_g only due to the Coulomb blockade charging

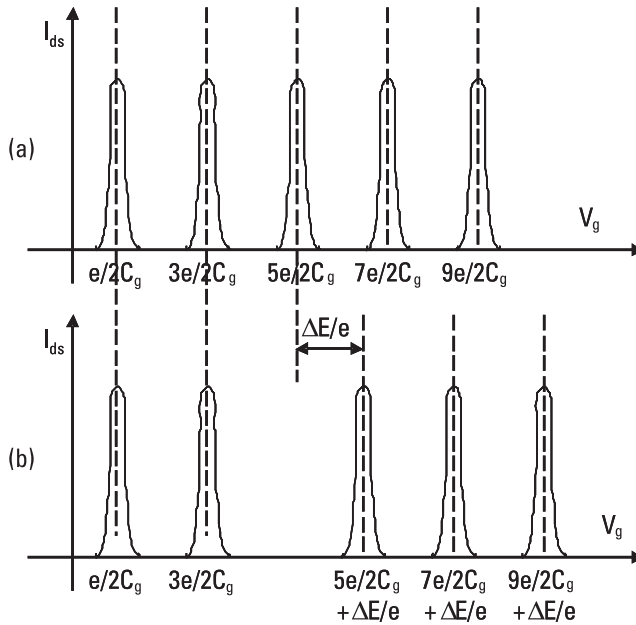


Figure 8.8 Comparison of V_g in a Coulomb oscillator (a) without and (b) with consideration of the quantum effect. When we take the quantum effect into consideration, the gate voltage shifts to higher direction by $\Delta E/e$.

energy until the second energy level is completely fulfilled and the ninth electron has to go to the third energy level.

As a matter of fact, this process is very much like filling the energy levels of atoms in the periodic table with two levels of quantum numbers; the first is the principal one that represents the energy level of quantum dots due to quantum mechanics and the second is equal spacing in the energy level from the excessive charging energy due to the Coulomb blockade. This is why quantum dots are sometimes called “artificial atoms” and a complete theoretical study of SETs includes not only electron–atom interactions like the one used in band theory, but also electron–electron interactions using “many-body theory.”

8.1.4 Fabrication of SETs

As we mentioned in Part II, as in the fabrication of all nanostructures, fabrication of single-electron transistors can also be divided into top-down and bottom-up approaches. Top-down approaches usually produce the Si channel first and then pattern it to the quantum dots (QD) level. The representative top-down approaches include pattern-dependent oxidization, direct write Si QD using FIB or electron beams, and so forth. The bottom-up approaches usually start

by making the semiconductor nanodots first. Here, we discuss these methods and also introduce some alternative methods based on self-assembly and phase separation or transformation.

8.1.4.1 Pattern-Dependent Oxidization (PADOX)

Takahashi et al. [3] first proposed this PADOX approach in 1995. As shown in Figure 8.9, a one-dimensional Si single wire was first patterned on SOI (Si on insulator) wafers. On each side of this wire, there are 2D large Si areas serving as the source and drain. Then the wafers were processed in a dry O_2 environment for high-temperature oxidization at 1000°C . As we all know, silicon oxidation is controlled by the oxygen diffusion step. Oxygen atoms can penetrate not only the upper surface of Si, but also the Si/SiO₂ interface. Therefore, the oxidation process occurs more on the pattern edge. As a result, this process not only decreases the width and height (or radius) of the Si nanowire, but it also constrains the two ends with thin dielectric layers and thus forms two tunnel junctions at both the source and drain sides for a single-electron transistor.

Other researchers have developed other versions of PADOX such as vertical PADOX in order to improve the throughput of fabrications as well as make multiple islands across the source and drain as shown in [4].

8.1.4.2 Patterned SET

Today with the advanced patterning techniques introduced in Part II, we can actually pattern Si quantum dots. For example, the focus ion beam (FIB) direct write technique has been used to fabricate SET [5]. In this case, a single silicon

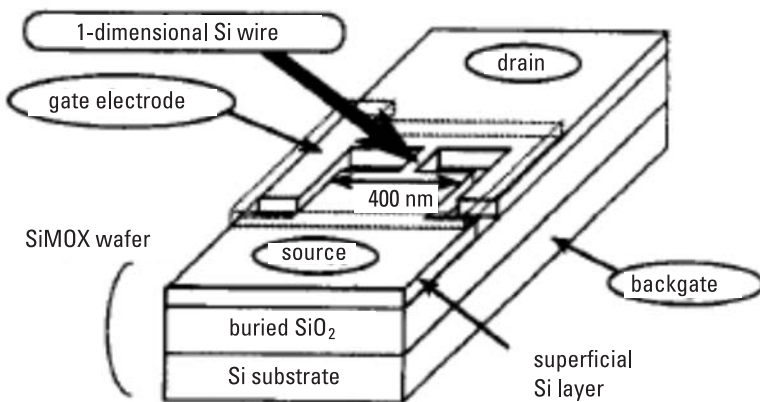


Figure 8.9 Schematic diagram of an Si SET. (Reprinted with permission from [3]. Copyright 1995 IEEE.)

quantum dot was produced between the source and drain directly using a FIB. However, due to the limitation in beam size, use of the FIB method results in a typical silicon quantum dot size of about 20 to 30 nm.

Another interesting method was introduced by Alivisatos's group in Berkeley [6]. They reported that they could pattern the source/drain area using standard optical lithography and then fabricate a nanowire between the source and drain using electron-beam lithography. After the nanowire is formed, a nanometer size gap as the tunneling junction between the source and drain was introduced through electromigration-induced wire breaking. However, due to the large current density passing through the nanowire, the resistance of the tunnel junction formed varied from 10 k Ω to 100 G Ω , indicating an unstable process control.

8.1.4.3 Nanoparticle-Based SET

For the bottom-up approach, we need to make the nanoparticle first. The methods used to make semiconductor nanoparticles are described in Part II. Many researchers [7–9] have reported on the fabrication of SETs based on this approach. For example, Klein et al. [7] reported on 5- to 10-nm CdSe nanocrystals produced by colloidal chemistry bounded between two Au plates as the source and drain electrodes. The bottom-up approach can easily bring quantum dot sizes down to the 5-nm region where both the Coulomb blockade and quantum effects will take effect. However, the disadvantage is that particle generation in large part is random so it is good for laboratory-scale studies, but hard to control for volume manufacturing.

8.1.4.4 Self-Assembly-Based SET

We mentioned earlier that the top-down approach usually is very controllable for industry-scale manufacturing, but the quantum wire/dot size is usually limited to 30 to 50 nm. With a bottom-up approach, we can easily make quantum dots with sizes down to 5 to 10 nm where the single-electron effect is more obvious but the process variation leads to highly random location, preventing its use for any practical applications other than research and exploration studies.

An alternative method introduced here is self-assembly as shown in Figure 8.10 [10]. This technique utilizes both a traditional top-down lithography approach and phase separation. The top-down lithography is designed to define the location of the quantum dots. The phase separation of a diblock polymer is used to create nanosize porous film, which then serves as a hard mask for either RIE (to etch Si) or deposition of amorphous Si. Either way silicon quantum dots are formed for SET applications.

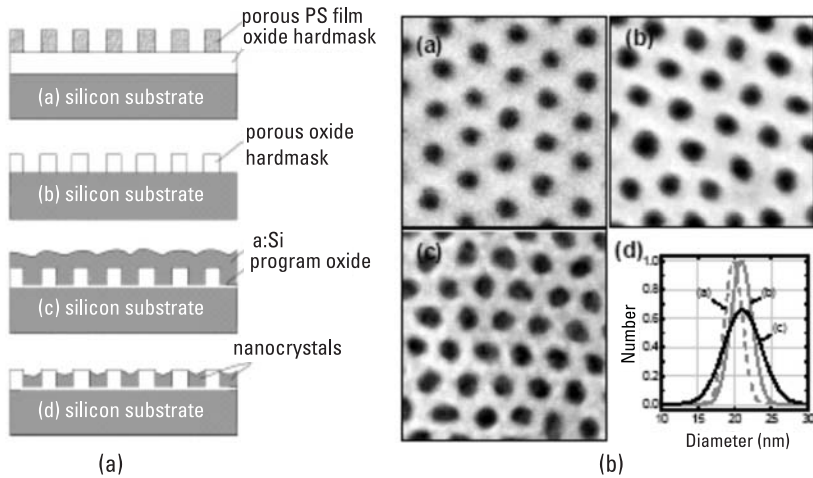


Figure 8.10 (a) Illustration of process flow that uses phase separation of diblock copolymer as the hard mask for fabricating an array of Si nanocrystals and (b) SEM images for a porous PS film hard mask, porous oxide hard mask, and array of Si nanocrystals. (Reprinted with permission from [10]. Copyright 2003 IEEE.)

8.1.4.5 Phase Transformation-Based SET

It has also been reported that self-formed Si nanocrystals can be produced from a Si-rich SiO_x substrate. According to its phase diagram, Si has no solubility at all in SiO_2 . So a Si-rich SiO_x film deposited by chemical vapor deposition (CVD) or PECVD will generate some Si nanocrystals (as shown in Figure 8.11) embedded in the original amorphous SiO_2 film after a high-temperature anneal. This is the most compatible method with today's IC manufacturing for generating nanosize quantum dots. However, the subsequent process for patterning those quantum dots into a single-electron transistor remains a challenge.

Other methods include spinodal decomposition-related phase transformation where the nanoparticle size is determined by the wavelength of the local compositional fluctuation. Because spinodal decomposition occurs at the instability point in the phase diagram, the transformation is simultaneous with no nucleation barriers. The key modulator is the degree of undercooling and misfit strain. There are reports that claim to use this method to produce nanocrystal Al or Au particles for single-electron transistors. (Figure 8.12) [11].

In this section, we introduced the basic concept of a single-electron transistor as well as a fundamental operating principle based on the Coulomb blockade and quantum effect. We also discussed several methods for fabricating single-electron transistors. The advantages of SETs are lower power consumption due to single-electron transportation and good scalability. One disadvantage is that, in most cases, it will have to operate at low temperature in order to avoid

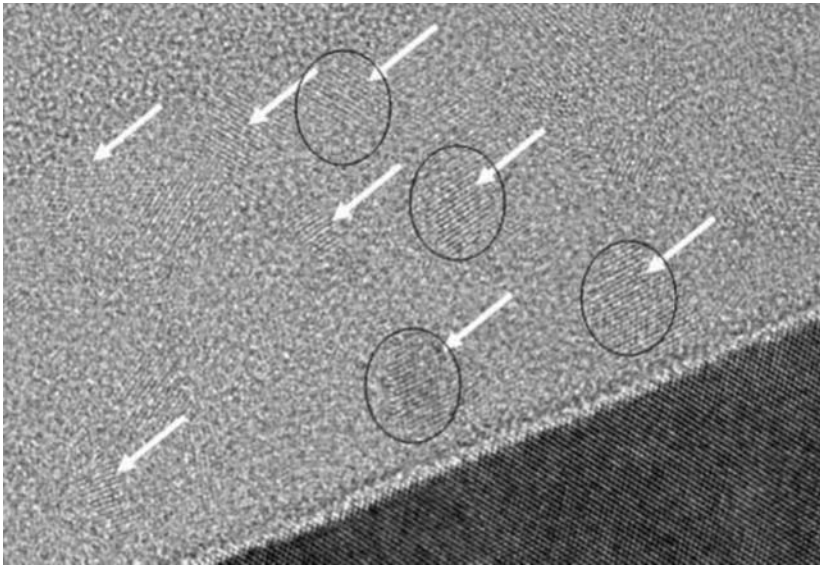


Figure 8.11 High-resolution TEM images showing Si nanoparticles embedded into SiO₂ film after phase separation from a Si-rich SiO₂ film deposited by PECVD methods.

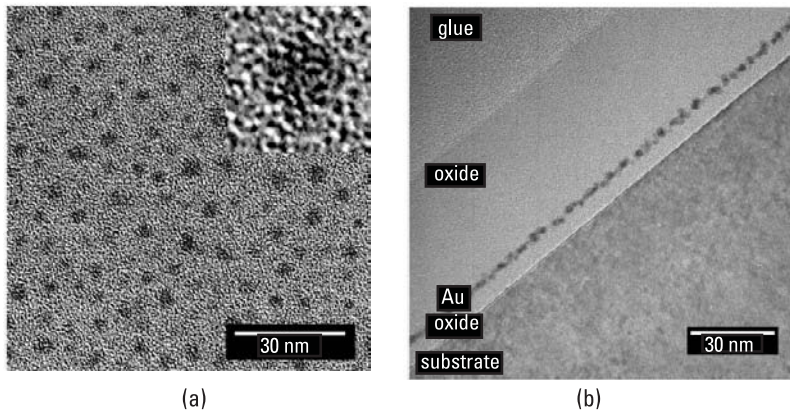


Figure 8.12 High-resolution TEM for (a) plan view and (b) cross-sectional view of postanneal Au₂₅Si₇₅ films showing an Au nanoparticle (3 nm) formed by spinodal decomposition. (Reprinted with permission from [11]. Copyright 2005 American Institute of Physics.)

fluctuation due to thermal energy. High output impedance is also an issue due to the intrinsic property of tunneling.

8.2 Carbon Nanotube-Based Nanoelectronic Device

Since their discovery by Iijima in 1991 [12], carbon nanotubes (CNTs) have been the most studied and also the most popular nanomaterials in the last decade due to their unique mechanical, electrical, and functional properties and relatively easy fabrication. In Chapter 3, we briefly introduced CNTs and their unique crystal structure including chiral angles. In this section we discuss the properties of CNTs first, then introduce several methods for fabricating CNTs, and end with their applications on nanoelectronic devices.

8.2.1 Introduction to CNTs

As we discussed in Chapter 3, each carbon atom has four electrons in the outer shell. They can form either sp^3 hybridization as in diamond, or sp^2 hybridization as in graphite and graphene. A carbon nanotube can be considered to be a rolled-up structure from a 2D graphene sheet. The chiral angle or vector (n, m) determines how the graphene sheet has been rolled up and thus largely the electronic properties of the CNT.

8.2.1.1 Band Structure of Graphene and Graphite

A 2D graphene sheet is like an extended version of the benzene ring discussed in Chapter 3. Each carbon atom forms three σ bonds through sp^2 hybridization with its three closest neighbors. The $2p_z$ orbit can be shared among neighboring atoms and thus form both bonding π and antibonding π^* bands through overlap among adjacent atoms. Based on this, scientists have derived a theoretical expression for the dispersion relationship between energy state E and wave vectors k_x and k_y of graphene as follows [13]:

$$E(k_x, k_y) = \pm \beta \left[1 + 4 \cos\left(\frac{\sqrt{3}}{2} k_x a\right) \cos\left(\frac{1}{2} k_y a\right) + 4 \cos^2\left(\frac{1}{2} k_y a\right) \right]^{1/2} \quad (8.13)$$

where β is the overlap integral as defined previously in Chapter 3, characterizing the overlap of the electron cloud between nearest neighbors; and a is the in-plane lattice constant and $a = 0.246$ nm for carbon. The two different signs are

denoted for the π and π^* bands. Figure 8.13 plots this band energy in the space of k_x and k_y . It is apparent that at six corners (of the Brillion zone) in k_x and k_y space, the π and π^* bands are touching each other, whereas in all other locations, they are separated by a bandgap. At absolute zero degree, the π band is completely occupied by electrons, whereas the π^* band is completely empty. However, at a certain specific point, the two bands just touch each other and the bandgap is zero. Therefore, graphene is called a “gapless” semiconductor. On the other hand, graphite is a set of graphite planes stacked together. Additional electron–electron interaction is formed between graphene planes, leading to an overlap between the π (valence) band and the π^* (conduction) band. Therefore, graphite is metallic and electrically conductive.

8.2.1.2 Band Structure of CNT

The band structure and dispersion relationship of graphene described earlier assume the infinite 2D plane. A CNT is a rolled-up form of graphene plane with chiral vectors $\mathbf{C}_h = (n, m)$. The axial direction along the long tube axis is macroscopic and by all practical means can be considered infinite, especially when compared with the radius/circumference direction. On the other hand, the circumference direction is only on the atomic scale and has a periodicity defined by chiral vectors \mathbf{C}_h . This periodic boundary condition of CNTs determines that only a discrete set of the (k_x, k_y) state is allowed in the dispersion

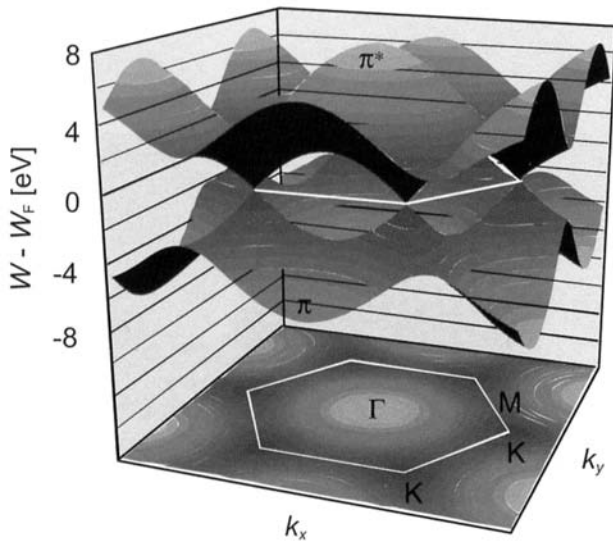


Figure 8.13 3D illustration of band structure of graphene in k_x, k_y space. (Reprinted with permission from [13]. Copyright 2005 Wiley-VCH Verlag GmbH & Co. KGaA.)

relationship in Figure 8.13. Using a similar periodic boundary condition as in the crystal structure, we obtain

$$\mathbf{C}_b \cdot \mathbf{k} = 2j\pi \quad (8.14)$$

where \mathbf{C}_b is the chiral vector, \mathbf{k} is the wave vector, and j is an integer. For example, for the armchair type configuration with $n = m$ in chiral vectors, the circumference direction (or \mathbf{C}_b vectors) is actually parallel to the hexagonal edge. Therefore, the \mathbf{k} direction for an armchair CNT is the same as the k_y direction in a graphene sheet:

$$\mathbf{k} = k_y = \frac{2j\pi}{|\mathbf{C}_b|} = \frac{2j\pi}{na\sqrt{3}} \quad (8.15)$$

Here, if we have a CNT of (3, 3), $n = 3$, j can take any integer between -3 and $+3$ and, thus, the allowed states are those lines labeled in Figure 8.14 [13]. If we use these lines to do a cross section of the graphene band structure, it is clear that for the armchair configuration, all six corners that are touching each other in the graphene band structure are the allowed state for CNT armchair configuration. Therefore, as shown in Figure 8.14(c), the armchair CNT is a gapless semiconductor. Furthermore, once we roll up the graphene sheet to the small dimensions of a CNT, the electron interaction is stronger, causing more overlap in π and π^* bands. Therefore, in reality, a CNT with an armchair configuration is metallic and conductive.

On the other hand, if we have a CNT with arbitrary chiral vectors, the \mathbf{k} vector calculated from (8.14) will also be arbitrary. Therefore, when plotted in Figure 8.14 and used as a cross section for the graphene band structure, there is no guarantee that this set of \mathbf{k} lines will pass those six corner points

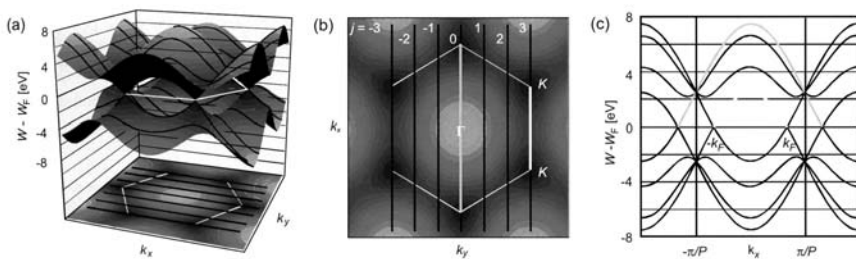


Figure 8.14 (a–c) Band structure of a CNT with the armchair configuration showing that it is metallic. (Reprinted with permission from [13]. Copyright 2005 Wiley-VCH Verlag GmbH & Co. KGaA.)

where the valence and conduction bands are touching. Therefore, in the general case, those CNTs are semiconductors due to the existence of the bandgap.

8.2.2 Fabrication of CNTs

Typical fabrication methods for CNTs include PVD methods such as electric-arc discharge or laser vaporization, and the CVD method, which sometimes is combined with other self-assembly or prepatterned structures.

8.2.2.1 PVD Methods

As we discussed in Chapter 5, the physical vapor deposition method involved a target that is made of the same materials as the desired film and an energy/momentum transfer mechanism that “sputters” off the target. Indeed, the first CNT was discovered by Iijima in 1991 [12] using one of the PVD methods called electric-arc discharge. This is similar to the method we introduced in Chapter 3 to fabricate the fullerene buckeye ball (C_{60}). In this method, two graphite electrodes are placed very close to each other inside a chamber that is full of inert gas. When a voltage is applied across the two electrodes, an electric-arc discharge occurs between the two electrodes, heating the electrode locally up to thousands of degrees and thus evaporating the carbon atoms from the electrode. The carbon atoms then recrystallize at the end of the negative electrode, forming a multiwall CNT (MWCNT).

Since Iijima first discovered MWCNTs using this method, it has been proven the easiest and the most reliable way to manufacture CNTs. Today, people can make CNTs in grams using this method. It is cheap and easy, but difficult to control. The CNTs produced using this method are usually multiwalled with a radius from 4 to 40 nm. It has been reported [14] that if we introduce some transition metal into the chamber as a catalyst, high yields of single-wall CNTs (SWCNTs) may be obtained.

The alternative method is laser-enhanced evaporation [15]. In this method, a graphite target is placed in the middle of a furnace at a high temperature of 1200°C and hit by a laser beam with a series of pulses. With every laser pulse, the carbon atoms are vaporized from the target surface. These atoms will continue to flow downstream in the vapor and eventually escape out of the furnace and recrystallize and grow to become CNTs on the cold finger outside the furnace. In this case, Co or Ni transition metal can be powdered into the graphite target and thus SWCNTs can be produced in large amounts.

8.2.2.2 CVD Methods

PVD methods can produce ropes of CNT, but CVD methods provide better control on CNT growth location and thus are more compatible with current Si manufacturing technology. In 1998, Hongjie Dai [16] and his group first

produced a single CNT using a CVD method. Since then several growth mechanisms have been proposed for CVD growth and it is still a debatable issue today in nanomaterials science.

First, the fact that a transition metal such as Co or Ni is always needed in order to form single-wall CNTs makes people think these metals serve as a catalyst. If this is the case, the catalyst will be left on the tip of the CNT. However, this is not observed. More recent developments in *in situ* TEM in which a SWCNT was grown by a CVD method inside a TEM chamber allowed direct observation of CNT growth [17]. Based on this observation, a vapor–liquid–solid (VLS) root-growth model was proposed and quickly accepted by many researchers [18]. As shown in Figure 8.15, in this model, transition metal nanoparticles still serve as a catalyst. Hydrocarbon is disassociated at the metal surface into H and C, which then dissolve into transition metals. When carbon becomes supersaturated inside the metal, it starts to precipitate out in the form of a curved-up graphene sheet and eventually leads to a fullerene cup. After that, more hydrocarbons are disassociated at the metal surface and then more carbon precipitates out and is inserted into the root of the fullerene cup, forming elongated carbon nanotubes. This is most popularly accepted growth mechanisms for CNT by CVD methods today.

The advantages of the CVD method is its compatibility with current microelectronics processes. For example, it is easy to control the growth location of CNTs by simply patterning the catalyst onto the wafer. A typical CVD method uses hydrocarbon C_nH_m such as methane as precursors for carbon [15–18]. Wafer/Si substrate is sitting in the middle of the deposition chamber that can be heated up to 1000°C . As we mentioned earlier, the catalyst can be

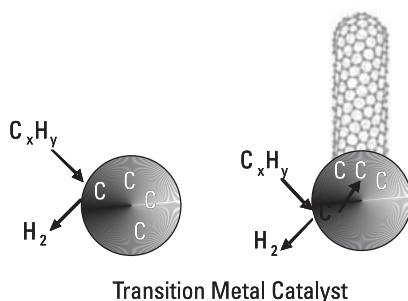


Figure 8.15 Illustration of VLS mechanism for the formation of an SWCNT from a transition metal particle as catalyst by the CVD process. First, hydrocarbon is decomposed at the surface of the transition metal nanoparticle. Carbon then dissolves into the metals. Once carbon becomes saturated, a carbon nanotube will be formed by incorporation of further carbon into the metal–carbon bonds at the root of the CNT.

prepatterned so CNT will grow only in the designated area. This is a very useful feature for CNT-based devices, which are discussed in the next section.

8.2.3 CNT-Based Devices

Because CNTs can be either metallic or semiconducting, they can be used as either interconnects such as vias or as FETs.

8.2.3.1 CNT Interconnect

As discussed earlier, a CNT with an armchair configuration is conductive due to the overlap between its valence and conduction bands. In addition, the transportation of electrons inside a CNT is actually scattering free and thus called *ballistic transport*. The theory behind this is related to the band structure and density of the states near the Fermi level, which is beyond the scope of this book. Interested readers can find further information in [13].

Due to the unique transport properties of CNTs, they have drawn much attention in recent years as interconnects, especially as vias and contacts, because of their morphology. As the device size continuously scales, the interconnect size also shrinks accordingly. In the most advanced technology nodes, the via and contact sizes are so small and the aspect ratio of height versus radius is so large that it poses a significant challenge for traditional tungsten CVD fill for contacts and Cu electroplating for vias. Additionally, due to the smaller size of vias and contacts, Cu vias have been plagued by electromigration issues with increasing current density. On the other hand, CNTs can be grown in situ inside the via/contact, in which case there is no intrinsic fill problem. This type of CNT is also immune to the electromigration issue due to its ballistic transport of electrons. Both make the CNT a perfect candidate for via interconnects.

This concept was first proposed by Nihei et al. [19] and has drawn much attention in recent years. For example, many researchers [19–21] have successfully grown CNTs on metal salicide, such as CoSi or NiSi as the contact to the source, drain and gate. In this case, both CoSi and NiSi can serve as catalysts to grow CNT based on the mechanisms discussed previously. However, making CNT vias in a Cu interconnect is more difficult because of the temperature constraints. The normal CVD growth method for CNTs that we discussed previously involves temperatures around 800° to 1000°C, much higher than the interconnected metal can endure. Therefore, the high-temperature CVD method cannot be directly used for back end of the line (BEOL) CNT interconnection unless we can drop the growth temperature to the 400° to 450°C range, which is determined by the Cu interconnect and low-*K* dielectrics. As shown in Figure 8.16, when enhanced by either plasma or microwaves, together with an Ni catalyst applied at the bottom of the via, the CVD process temperature

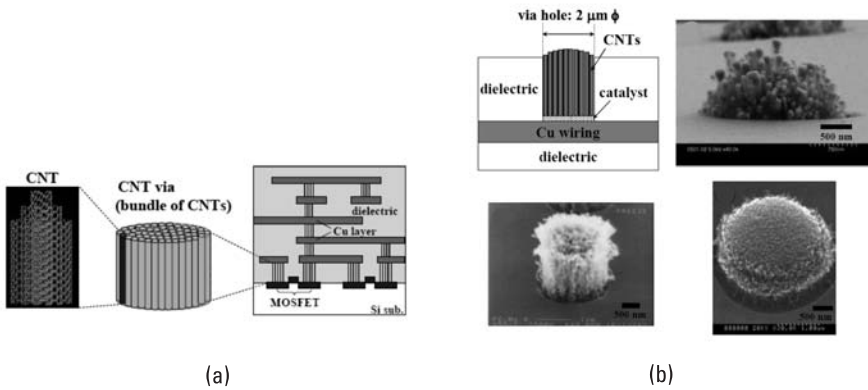


Figure 8.16 (a) Illustration of CNT via device and (b) SEM images showing vertically aligned MWCNT bundle in via hole by different CVD methods for future VLSI interconnect. (Reprinted with permission from [21]. Copyright 2004 IEEE.)

can be lowered to about 400°C, making a CNT via interconnect possible [19, 20].

The electrical data revealed that the current–voltage relationship is actually ohmic, indicating a good contact between the CNT and Cu metal lines [21]. Obviously, most of the CNT in this case should be metallic with the armchair configuration. In fact, the CNTs obtained through this method are usually multiwalled ones, which can also help conductivity due to there being more than one wall/channel per tube.

8.2.3.2 CNT Transistor

A traditional MOSFET is operated based on the accumulation, depletion, and inversion of silicon materials. When the gate voltage is high enough to turn the transistor on, electrons flow from the source to the drain through an inverted Si channel. In this case, the conductive channel of inverted silicon can be considered to be a 2D electron gas because the channel depth is controlled by the gate voltage. With the scaling of the device, the channel length also becomes smaller. However, the channel width is still very large compared to the other two dimensions. Much effort has been put into this front to simultaneously shrink both the channel length and width direction for the next generation of MOSFETs, such as 3D gates and FINFET. However, if a CNT is used as the channel between the source and the drain in a similar field emission transistor system (sometimes called a CNT-MOSFET), the electron system is one dimensional by definition due to the morphology of the CNT. In addition, due to the ballistic transportation properties of CNT, the carrier mobility is much higher inside the CNT than the silicon. Therefore, a much higher drive current is expected for the CNT-MOSFET compared to the traditional Si-MOSFET.

In 1998, Tans et al. [22] from Delft University were able to successfully fabricate a carbon nanotube-based MOSFET device. In their research, a single CNT was placed between a platinum source and the drain as the active channel. The backside silicon acted as a gate to turn this transistor on and off. The characteristic I–V curve of this transistor is very similar to that of a traditional Si-based MOSFET. This study is landmark research for nanomaterials because, for the first time, researchers have been able to successfully produce a nanoscale device that has the equivalent of or better performance than a traditional microelectronic device at room temperature and normal operating conditions. This really represents the beginning of the nanoelectronics era. Since then, many developments have been reported in the literature about successful fabrications of CNT-MOSFETs [23–25].

To successfully fabricate CNT-MOSFETs, the following challenges have to be overcome: First, the CNT has to be a semiconductor in order to serve as the channel. As we know, the normal distribution of a CNT includes both metallic and semiconductive materials. Moreover, almost all multiwall CNTs are conductive due to the many paths available for electrons. Therefore, it is critical to grow a single-wall CNT with the semiconductive properties of a MOSFET channel. Kong et al. [16] from Stanford University has developed a special CH₄-based CVD process to produce SWCNTs only. This will greatly reduce the conductivity of the channel during the “off” state. However, the bundle of SWCNTs between the source and the drain can still either be metallic (armchair configuration) or semiconductive (all other configurations). So, to filter out the metallic CNT, a large current was applied between the source and the drain while the transistor is off to “burn off” those metallic carbon nanotubes. Therefore, only semiconductive SWCNTs are left as the channel for the CNT-MOSFET.

The second challenge is to make MOSFET structures. This includes figuring out how to place a CNT between the source and the drain and how to integrate the gate around the channel. Today, the easiest and effective way is first to pattern the source and drain electrodes (Au, Al, or Pt) on a Si wafer and then grow a high-density CNT randomly so that there will be some amount of CNT between the source and the drain as channels. The gate can be either at the backside as in the original research done by Tans et al. [22] or on the side of the CNT channel, which is unlikely. Most recently, Javey et al. [26] have developed a top gate structure for CNT-MOSFETs in which they use ALD (see Chapter 5) to deposit an 8-nm high-*K* dielectric (ZrO₂) in order to wrap around the CNT without any damage or degradation to it. The high-*K* insulator not only serves as a gate oxide but also protects the CNT before the gate materials are deposited. This top gate structure has achieved a 70 mV/decade subthreshold swing, which almost approaches the theoretical limit for a MOSFET.

8.3 Spintronics

Spintronics, as it is called, is a new type of device that is based on the spin properties of electrons. In this section, we briefly explain the fundamental principles of spintronics and then discuss the methods for generating, transporting, and detecting spin-polarized current before we introduce several different types of spintronics devices and their applications.

8.3.1 Operating Principles of Spintronics

As introduced in Chapter 1, an electron has two spin states, one spin up and the other spin down. This is an intrinsic property of electrons. The operating principle of spintronics is to generate, transport, and detect the current of an electron with specific spin states, which, in turn, relies fundamentally on the spin properties of electrons and their interactions with the external magnetic field.

8.3.1.1 Spin Magnetic Momentum

The spin of an electron generates a magnetic momentum that can interact with the external magnetic field. The spin magnetic momentum of a single free electron μ_s can be described as a function of spin \mathbf{S} :

$$\mu_s = -g\mu_B \mathbf{S} \quad (8.16)$$

where μ_B is the Bohr magneton and g is the dimensionless factor for the electron. For a free electron it is about 2. Because the electron's spin \mathbf{S} is $1/2$, the spin momentum of a free electron is approximately the Bohr magneton. However, g is generally different in a solid system compared with the free electron value. Furthermore, the spin momentum of electrons in a solid will also be the net sum of the spin magnetic momentum from an assembly of electrons.

Because spin generates magnetic momentum, it will interact with the external magnetic field. When applied as an external magnetic field (B), the total energy will have an extra term that can be expressed as $E = E_0 + E_M = E_0 - \mu \cdot B$. This is called the Zeeman effect. Therefore, if the direction of spin magnetic momentum is aligned with the external magnetic field, the overall system energy will decrease and vice versa.

8.3.1.2 Giant Magnetoresistance Effect

Discovered in the 19th century, the classic magnetoresistance theory stated that a material's resistance will change when an external magnetic field is applied. This is based purely on classical physics, which said that the carrier velocity

and mobility would change due to the existence of the Lorenz force introduced by the external magnetic field.

However, in 1988, French scientist Albert Fert and German scientist Peter Grünberg simultaneously discovered the giant magnetoresistance (GMR) effect, which led to their sharing the Nobel Prize in Physics in 2007 [27, 28]. They stated that in a trilayer, sandwiched structure consisting of ferromagnetic–nonmagnetic–ferromagnetic thin films (shown in Figure 8.17), the magnetic momentum directions in two ferromagnetic thin films will be antiparallel due to weak coupling across nonmagnetic spacer film if there is no external magnetic field. Thus, the electrical resistance will be high. On the other hand, if an external magnetic field is applied, the direction of magnetization in the two ferromagnetic thin films will be parallel and thus the electrical resistance will be significantly reduced. This is the giant magnetoresistance effect.

As we mentioned in a previous section, the spin magnetic momentum turns to align with an external magnetic field in order to minimize the total energy of a system. In general, a current is spin neutral, meaning that statistically the net spin of all electrons is about zero. However, when electrons pass through the first ferromagnetic films as shown in Figure 8.17, the directions of spin magnetic momentum are forced to align with the direction of the magnetization of the film, thus producing a spin-polarized current. When this current enters the second ferromagnetic film with antiparallel magnetization to the first film

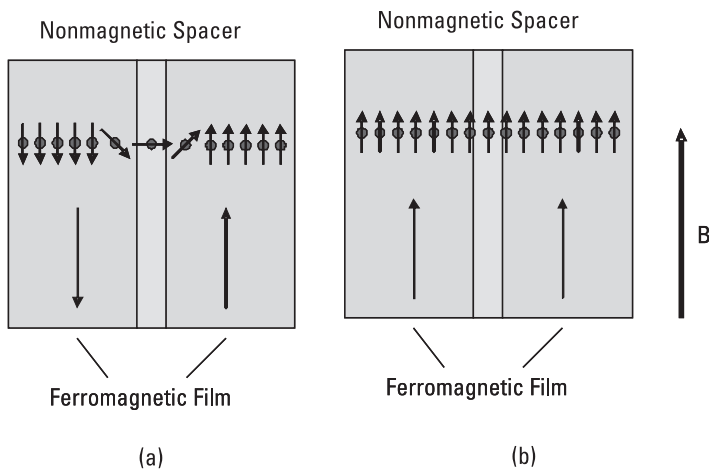


Figure 8.17 Illustration of GMR effect on a ferromagnetic–nonmagnetic–ferromagnetic trilayer sandwiched structure. (a) When no external magnetic field exists, two ferromagnetic films will be antiparallel, forcing a spin change during electron transportation and thus creating high resistance. (b) When an external magnetic field exists, the two ferromagnetic films will be parallel in magnetization and thus electrons can be transported through without a magnetic momentum change.

due to the absence of the external magnetic field, the spin magnetic momentum of the carriers in the current is forced to flip again, leading to magneto scattering and thus high electrical resistance. On the other hand, when an external magnetic field is applied, the magnetization of both ferromagnetic films becomes parallel and thus the spin momentum can remain the same while passing the second film—this can greatly reduce the electrical resistance and leads to the GMR effect.

8.3.2 Spintronics Devices

We now know that the operating principles of spintronics are based on the quantum effect of spin magnetic momentum and GMR. Hence, a spintronics device should be able to manipulate the spin of electrons or, more specifically, generate, transport, and finally detect the spin-polarized current.

A typical spintronics device consists of at least two basic elements—a spin battery or pump that can generate spin-polarized current and a sensor that can detect the spin polarization. In some cases, there is also a channel between the battery and the sensor that can inject, transport, or even change the spin-polarized current without much relaxation.

In this section, we discuss several different types of spintronics device with an emphasis on spintronics-based field-emitted transistors.

8.3.2.1 GMR Storage Device

The most common use of a GMR device is as a hard-disk drive head that can read and write information from a hard-disk drive (HDD). Traditionally an HDD head reads the data from the hard disk by sensing the rate of change of magnetisms through small coils inside the head. Lately with the discovery and development of GMR, spintronics-based devices have begun to be widely used in the data storage industries as read heads (the write head usually is a thin-film inductive head that can generate magnetic field). The application of a GMR-based device in HDDs enables aggressive capacity doubling every 12 months. As of mid-2008, the highest capacity for a commercially available hard drive was 1.5 TB (Seagate Barracuda).

A typical read head sensor also consists of three sandwiched layers similar to the structure shown in Figure 8.17. The difference is that the thickness of the nonferromagnetic spacer between the two ferromagnetic materials is such that the exchange coupling between the two ferromagnetic films is at a minimum. As a result, it is possible to independently control and switch the magnetization of the two ferromagnetic films. As discussed earlier, if they are antiparallel, the resistance will be higher than if they are parallel. This device is called a spin valve.

In the read head that assembles the spin valve, one ferromagnetic thin film has a fixed (also called pinned) magnetization, while the other film has a

free one. When the sensor is very close to the hard disk with a small air gap, the passing magnetic bit in the hard-disk area (the bit stored as 1) under the head will change the magnetization of free layers and thus the magnetoresistance of the film. Therefore, if a current passes through the read sensor, the change due to GMR will be detected and a “1” will be read from the hard disk. On the other hand, if data stored in the hard disk is zero—that is, there is no net magnetic field passing under the head, the resistance will not be high, and the current can flow with no change—this is the state of “0.” This is the basic operating principle for a GMR-based read sensor in an HDD head. More recent developments have been made on the tunneling effect and other improvements that have enabled further expansion of capacity [29].

8.3.2.2 Spin Field Emission Transistor (SFET)

The spin field emission transistor was first proposed by Datta and Das in 1990 [30]. This model, which later become the prototype spintronics device, is based on spin injection and detection. As shown in Figure 8.18, two ferromagnetic layers with parallel magnetization act as the source and drain, respectively. When the gate voltage is zero, the spin-polarized current that is generated in the ferromagnetic source is injected into the channel and transported to the drain region with its spin preserved (that is, minimum relaxation on spin polarization). Because the ferromagnetic layers in the drain side share a parallel magnetization with the source, the magnetoresistance is low and, therefore, the current flies through. This is the transistor “on” state. On the other hand, if a gate voltage is applied, the electric field in the channel region forces electrons to flip their spin direction from parallel to antiparallel to the drain side ferromagnetic films.

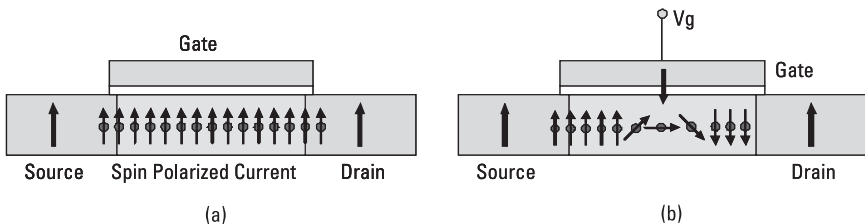


Figure 8.18 The Datta-Das SFET consists of a source and drain region with parallel ferromagnetic domains. (a) At zero gate voltage, the electrons’ spin state is preserved in the transport channel, enabling spin-polarized current to flow from the source to the drain; that is, the transistor is in the “on” state. (b) With an applied gate voltage, electrons change their spin state from parallel to antiparallel to the direction of magnetization of the ferromagnetic layer during transportation in the channel. This offers high resistance to the flow of current. Therefore, electron scattering occurs at the drain and no current flows from the source to drain; that is, the transistor is in the “off” state.

Therefore, the magnetoresistance for the channel current becomes very large and the transistor is in the “off” state.

Although only a model, the SFET has drawn significant attention in recent years [31–33] due to its potential low-power consumption in the “off” state and high current conductance in the “on” state. However, SFETs have at least these few main challenges:

1. *Spin injection*: To inject a spin-polarized current from the source into the channel without relaxation is not an easy task—the injection has to pass through the magnetic tunneling junction and p - n or Schottky junction.
2. *Spin transportation through the channel without much relaxation in an “on” state*: This means ballistic transportation for spin-polarized current through the channel without disturbance from temperature fluctuations or an undesired internal electric field.
3. *Effective gate control that can turn the transistor off completely*: Also the drain side junction will play a key role in stopping the current flow.

Much current research into materials design has been applied to this area [34, 35], including ferromagnetic and ferrosemiconductor materials. In addition, researchers have proposed many new different types of SFETs. (Interested readers can find some reviews in [31–33].) But the field still has a long way to go before it can overcome the challenges just discussed and replace traditional Si-based MOSFETs.

8.3.2.3 Magnetoresistive Random Access Memory (MRAM)

Random access memory (RAM) is frequently used in all kinds of electronics device. Including SRAM (static RAM) and DRAM (dynamic RAM), this type of data storage is mainly based on electric charge stored in a capacitor-like device and thus is volatile, meaning that when the external voltage is removed, the stored data will disappear. MRAM (magnetoresistive RAM), on the other hand, stores data not by an electric charge but by magnetic momentum in ferromagnetic films and thus it is nonvolatile.

First introduced by then Motorola and later Freescale [36], the operating principle of MRAM is still based on the GMR effect. As shown in Figure 8.19, a typical MRAM cell consists of one metal line above (called the bit line) and one metal line below (called the word line) a trilayer sandwich structure including a top and bottom ferromagnetic layer separated by an insulating layer. In a writing operation, the word line and bit line both have large current passing through and thus define a specific cell to write (this is why it is called RAM). Because programming (writing) current is large, it will generate an external

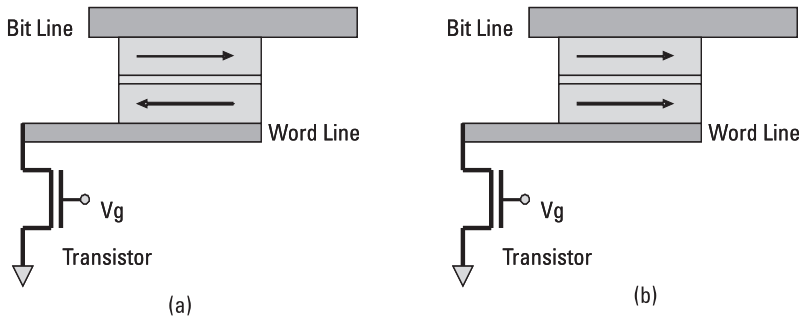


Figure 8.19 Schematic illustration of MRAM based on spintronics: (a) The magnetization of two ferromagnetic films is antiparallel; hence, no current is allowed to pass. This is the state of “0.” (b) The magnetization of the two ferromagnetic films is parallel. Resistance is small based on the GMR effect, so current is allowed to pass. This is the state of “1.” The magnetization of domains in the films is produced by passing a large current through the bit line and word line during programming.

magnetic field and thus align the directions of magnetization in two ferromagnetic plates. If two directions are parallel, it will be the state of “1” and if antiparallel, it will be the state of “0.” During the operation of reading, a relatively small current will pass through the memory cell to sense the polarity of the magnetization. If the directions from the two ferromagnetic plates are parallel, the magnetoresistance will be low and, thus, the current will pass through and be sensed by the transistor—it reads this bit as “1.” On the other hand, if they are antiparallel, the current cannot pass through due to the GMR effect and thus it reads this bit as “0.”

The advantages of MRAM are very fast read/write speeds and random access just as with SRAM. In addition, however, it is nonvolatile because the data storage is based on magnetization of ferromagnetic materials and the information will not get lost even after the external voltage is removed. The disadvantages of this type of memory are the relatively large cell size required and the high write current required in order to generate a magnetic field for programming.

A new approach called spin torque transfer (STT) has been developed in order to decrease the writing current [37–39]. This approach is actually based on the reverse effect of GMR. In the GMR effect, spin-polarized current will be detected and modulated by the ferromagnetic materials. In the STT approach, however, the spin-polarized current will actually transfer the magnetic momentum (or torque) to the ferromagnetic materials (a soft magnetic material will be used in this case) and thus modulate the magnetization of the ferromagnetic material itself. This technique has significantly lowered the amount of current required for MRAM programming and thus made it a potential candidate for

future data storage. Today, Toshiba and other vendors have successfully produced MRAM chips having up to a gigabyte capacity [40].

References

- [1] Moore, G. E., "Cramming More Components onto Integrated Circuits," *Electronics*, Vol. 38, No. 8, 1965.
- [2] Dubash, M., "Moore's Law Is Dead, Says Gordon Moore," Techworld, interview with Gordon Moore, April 13, 2005, available at <http://www.techworld.com/opsys/news/index.cfm?newsid=3477>.
- [3] Takahashi, M., et al., "Fabrication Technique for Si Single Electron Transistor Operating at Room Temperature," *Electronics Letters*, Vol. 31, No. 2, 1995, pp. 136–137.
- [4] Ono, Y., et al., "Fabrication Method for IC-Oriented Si Twin Island Single Electron Transistor," *IEEE International Electron Devices Meeting 1998 (IEDM '98) Technical Digest*, 1998, pp. 123–126.
- [5] Kim, T. W., et al., "Single Electron Transistor Operating at Room Temperature, Fabricated Utilizing Nanocrystals Created by Focused-Ion Beam," *Applied Physics Letters*, Vol. 80, No. 12, 2002, pp. 2168–2170.
- [6] Park, H. K., et al., "Fabrication of Metallic Electrodes with Nanometer Separation by Electro Migration," *Applied Physics Letter*, Vol. 75, No. 2, 1999, pp. 301–303.
- [7] Klein, D. L., et al., "A Single Electron Transistor Made from a Cadmium Selenide Nanocrystal," *Nature*, Vol. 389, 1997, pp. 1–3.
- [8] Ralph, D. C., C. T. Black, and M. Tinkham, "Gate Voltage Studies of Discrete Electronic States in Aluminum Nanoparticles," *Physics Review Letter*, Vol. 78, 1997, pp. 4087–4090.
- [9] Feldheim, D. L., and C. D. Keating, "Self-Assembly of Single Electron Transistors and Related Devices," *Chemical Society Reviews*, Vol. 27, 1998, pp. 1–12.
- [10] Guarini, K. W., et al., "Low Voltage, Scalable Nanocrystal Flash Memory Fabricated by Templated Self Assembly," *IEEE International Electron Devices Meeting 2003 (IEDM '03) Technical Digest*, 2003, pp. 22.2.1–22.2.4.
- [11] Chandra, A., and B. M. Clemens, "Gold Nanoparticles via Alloy Decomposition and Their Application to Nonvolatile Memory," *Applied Physics Letters*, Vol. 87, 2005, 253113-3.
- [12] Iijima, S., "Helical Microtubules of Graphitic Carbon," *Nature*, Vol. 354, 1991, pp. 56–58.
- [13] Waser, R. (Ed.), *Nanoelectronics and Information Technology*, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2005, p. 476.
- [14] Minea, T. M., et al., "Single Chamber PVD/PECVD Process for In Situ Control of the Catalyst Activity on Carbon Nanotube Growth," *Surface & Coating Technology*, Vol. 200, 2005, pp. 1101–1105.
- [15] Yudasaka, M., et al., "Single-Wall Carbon Nanotube Formation by Laser Ablation Using Double-Targets of Carbon and Metal," *Chemical Physics Letters*, Vol. 278, 1997, pp. 102–106.

-
- [16] Kong, J., A. M. Cassell, and H. Dai, "Chemical Vapor Deposition of Methane for Single-Walled Carbon Nanotubes," *Chemical Physics Letter*, Vol. 292, 1998, pp. 567–574.
 - [17] Sharma, R., and Z. Iqbal, "In Situ Observations of Carbon Nanotube Formation Using Environmental Transmission Electron Microscopy," *Applied Physics Letters*, Vol. 84, No. 6, 2004, pp. 990–992.
 - [18] Kukovitsky, E. F., S. G. L'vov, and N. A. Sainov, "VLS-Growth of Carbon Nanotubes from the Vapor," *Chemical Physics Letters*, Vol. 317, 2000, pp. 65–70.
 - [19] Nihei, M., et al., "Carbon Nanotube via Technologies for Future LSI Interconnects," *Proc. of the IEEE 2004 International Interconnect Technology Conference*, 2004, pp. 251–253.
 - [20] Graham, A. P., et al. "Towards the Integration of Carbon Nanotubes in Microelectronics," *Diamond and Related Materials*, Vol. 13, 2004, pp. 1296–1300.
 - [21] Nihei, M., et al., "Simultaneous Formation of Multiwall Carbon Nanotubes and Their End-Bonded Ohmic Contacts to Ti Electrodes for Future ULSI Interconnects," *Japan J. of Applied Physics*, Vol. 43, 2004, pp. 1856–1859.
 - [22] Tans, S. J., A. R. M. Verschueren, and C. Dekker, "Room Temperature Transistor Based on a Single Carbon Nanotube," *Nature*, Vol. 393, 1998, pp. 49–52.
 - [23] Li, S., et al., "Carbon Nanotube Transistor Operation at 2.6 GHz," *Nano Letters*, Vol. 4, No. 4, 2004, pp. 753–756.
 - [24] Javey, A., et al., "Carbon Nanotube Transistor Arrays for Multistage Complementary Logic and Ring Oscillators," *Nano Letters*, Vol. 2, No. 9, 2002 pp. 929–932.
 - [25] Derycke, V., et al., "Controlling Doping and Carrier Injection in Carbon Nanotube Transistors," *Applied Physics Letters*, Vol. 80, No. 15, 2002, pp. 2773–2775.
 - [26] Javey, A., et al., "High Dielectrics for Advanced Carbon-Nanotube Transistors and Logic Gates," *Natural Materials*, Vol. 1, 2002 pp. 241–246.
 - [27] Baibich, M. N., et al., "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Physical Review Letters*, Vol. 61, 1988, p. 2472.
 - [28] Grunberg, P., "Layered Magnetic Structures: History, Highlights, Applications," *Physics Today*, May 2001, pp. 31–37.
 - [29] Parkin, S. S. P., "Giant Tunnelling Magnetoresistance at Room Temperature with MgO (100) Tunnel Barriers," *Natural Materials*, Vol. 3, 2004, pp. 862–867.
 - [30] Datta, S., and B. Das, "Electronic Analog of the Electro-Optic Modulator," *Applied Physics Letters*, Vol. 56, No. 7, 1990, pp. 665–667.
 - [31] Zutic, I., J. Fabian, and S. D. Sarma, "Spintronics: Fundamentals and Applications," *Reviews of Modern Physics*, Vol. 76, 2004, pp. 323–410.
 - [32] Wang, B., J. Wang, and H. Guo, "Quantum Spin Field Effect Transistor," *Physical Review B*, Vol. 67, 2003.
 - [33] Wolf, S. A., et al., "Spintronics: A Spin-Based Electronics Vision for the Future," *Science*, Vol. 294, 2001. pp. 1488–1495.
 - [34] Silsbee, R. H., "Theory of the Detection of Current-Induced Spin Polarization in a Two-Dimensional Electron Gas," *Physical Review B*, Vol. 63, 2001.

- [35] Vazquez, M., et al., "Magnetic Nanoparticles: Synthesis, Ordering and Properties," *Physica B: Condensed Matter*, Vol. 354, 2004, pp. 71–79.
- [36] Tehrani, S., et al., "Progress and Outlook for MRAM Technology," *IEEE Trans. on Magnetics*, Vol. 35, No. 5, 1999, pp. 2814–2819.
- [37] Sun, J. Z., et al., "Spin-Torque Transfer in Batch-Fabricated Spin-Valve Magnetic Nanojunctions," *J. of Applied Physics*, Vol. 93, No. 10, 2003, pp. 6859–6863.
- [38] Hosomi, M., et al., "A Novel Nonvolatile Memory with Spin Torque Transfer Magnetization Switching: Spin-Ram," *IEEE International Electron Devices Meeting 2005 (IEDM '05)*, 2005, pp. 459–462.
- [39] Li, Z., and S. Zhang, "Magnetization Dynamics with a Spin-Transfer Torque," *Physical Review B*, Vol. 68, 2003.
- [40] Toshiba, "Toshiba Develops New MRAM Device Which Opens the Way to Giga-bits Capacity," November 6, 2007, available at http://www.toshiba.co.jp/about/press/2007_11/pr0601.htm.

9

Nano Biomaterials

The biological and physical sciences share a common interest in small structures. To the physical sciences, “nano” offers quantum phenomena and remarkable physical properties. Biology adds incredibly sophisticated nanomachines, operating by entirely classical molecular mechanisms. To the biological sciences, “nano” offers new tools—many from the physical sciences—that will be necessary to put together a conceptual model of life, and a fresh framework on which to hang ideas about functional aggregates. In this chapter, we present some ideas about interactions between nanotechnology and biology.

9.1 Introduction

A vigorous dialogue across the borders of these areas of science is developing around new materials and tools (largely from the physical sciences) and new phenomena (largely from the biological sciences). The physical sciences offer tools for synthesis and fabrication of devices for measuring the characteristics of cells and subcellular components, and of materials useful in cell and molecular biology; biology offers a window into the most sophisticated collection of functional nanostructures that exists.

There are important applications of nanoscience in biology and biotechnology; biology also provides unparalleled examples of functional nanostructures to excite the imagination of nanotechnologists of all persuasions. The story is, however, not entirely “nano”; it also includes structures having a wide range of sizes. When small structures are considered for biological applications, or when small biologically derived structures are determined to have remarkable properties, the size of the system can be “nano” but also “micro” [1]. Nature

has been a source of inspiration to the fabrication and manufacturing community by presenting biological components with structures having incredible functions. We focus on several examples of biomimetic nanotechnology in the next section. Following this, in Sections 9.3 through 9.5, we discuss some applications of micro/nanostructures as tools for biotechnology.

9.2 Biomimetic Nanotechnology

Biological materials and systems exhibit complex functionality and are composed of nanoscale components. Nature combines hard and soft materials, often in hierarchical architectures, to get synergistic, optimized properties and combinations of properties. Learning from nature is an important way for human science and technology to advance.

9.2.1 DNA Nanotechnology

A key property of biological nanostructures is molecular recognition, leading to self-assembly and the templating of atomic and molecular structures. For example, it is well known that two complementary strands of DNA will pair to form a double helix. DNA illustrates two features of self-assembly. The molecules have a strong affinity for each other and they form a predictable structure when they associate. Those who wish to create defined nanostructures would like to develop systems that emulate this behavior. Thus, rather than milling down from the macroscopic level, using tools of greater and greater precision (and probably cost), they would like to build nanoconstructs from the bottom up, starting with chemical systems.

To date, the most successful biomimetic component used for self-assembly has been DNA itself. Linear DNA double helices seem to be of limited utility, but one can design synthetic molecules that form stable branched structures, leading to greater structural complexity. Branched DNA molecules can be combined by “sticky-ended” cohesion [2, 3], as shown in Figure 9.1. In synthetic systems, sticky ends may be programmed with a large diversity; N -nucleotide sticky ends lead to 4^N possible different sequences. Sticky ends of sufficient length cohere by base pairing alone, but they can be ligated to covalency. Sticky ends form classic *B*-DNA (right-handed typical form of double helix DNA) when they cohere; thus, in addition to the affinity inherent in complementarity, sticky ends also lead to structural predictability. If the positions of the atoms on one component are known near the sticky end, the atoms of the other component are also known. This situation is usefully contrasted with, say, an antibody and its antigen. Although the antigen-combining site may be known, the orientation of the antigen within it cannot be predicted; it must be determined experimentally in each case.

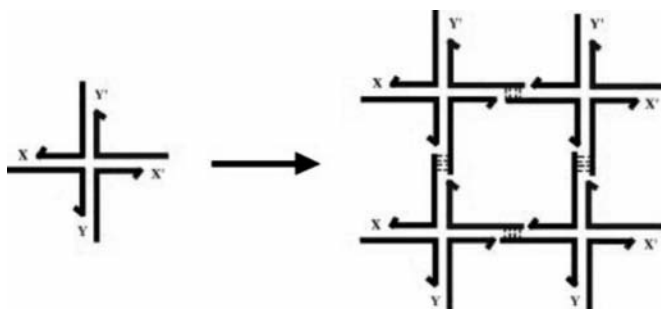


Figure 9.1 Formation of a 2D lattice from a junction with sticky ends. The X and Y represent sticky ends and X' and Y' are their complements. Four of the monomers on the left are complexed to yield the structure on the right. DNA ligase can close the gaps left in the complex, which can be extended by the addition of more monomers. (Reprinted with permission from [2]. Copyright 2001 American Chemical Society.)

The key static aims of DNA nanotechnology are to use DNA as scaffolding to crystallize biological macromolecules artificially for crystallography and to organize the components of nanoelectronics. The first, and likely the second, of these applications entails the assembly of DNA into periodic networks. Thus, the quadrilateral of Figure 9.1 would be most useful if extended to form a 2D or 3D lattice. The branched junctions shown in Figure 9.1 are not rigid enough to use as building blocks for a lattice. This problem has been solved by combining two branched junctions to produce DNA double-crossover (DX) molecules [4], which consist of two double helices fused by strands that cross between them to tie them together.

The control over DNA systems seems to be relatively robust. However, to what ends can it be used? The physical properties of DNA are important for understanding biological systems, but their utility in nanoelectronic devices is unproven. To get maximum use from the organizational capabilities of DNA, it will be necessary to combine DNA with other nanoscale systems, particularly inorganic systems, whose physical properties lend themselves to direct applications. These materials include inorganic nanocrystals and carbon nanotubes, which represent the most exciting potential species to organize in two or three dimensions. The large sizes and abundant functional groups on DNA tiles suggest that multiple functionalities could be attached to a single tile.

Recently, researchers have reported on a group of complex 3D geometric architectures of Au nanoparticles created using DNA tile-mediated self-assembly [5]. These are tubular nanostructures with various conformations and chiralities resembling those of CNTs. The NP tube assembly can be engineered both by the underlying DNA tile scaffolds and the nanoparticles themselves. The intrinsic dimensional anisotropy of the DNA tiles plays an important role in the kinetic control of the tube growth.

An array system formed from four DX DNA tiles was used [Figures 9.2(a, b)]. In the first design, two out of the four component tiles are modified. The central strand in the A tile was conjugated with a thiol group, which was then linked to a 5-nm Au NP in a 1:1 ratio so that, when self-assembled, each A tile carried a Au NP on one side of the tile (shown as the top side). The C tile was modified with a DNA stem loop extending out of the tile surface toward the bottom side [Figures 9.2(a, b)]. As illustrated in Figure 9.2(c), these four tiles were designed to self-assemble into a 2D array through sticky-end associa-

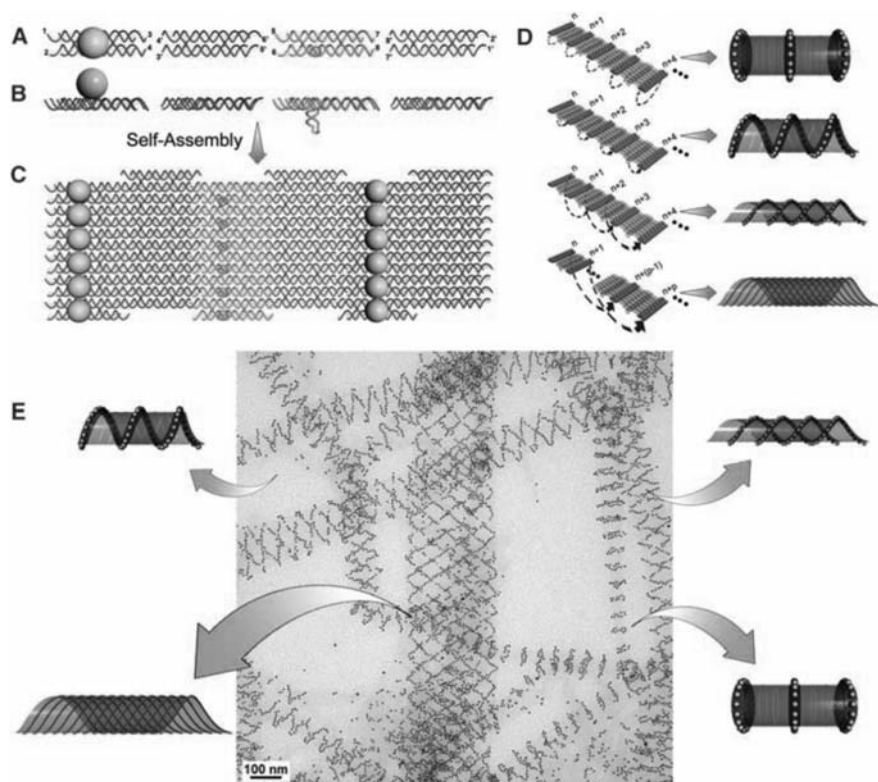


Figure 9.2 The design of a DNA tile system for the formation of a variety of tubular structures carrying 5-nm Au NPs. (a, b) Top and side views of the four DX tiles. The A tile carries a 5-nm Au NP on the top of the tile. The C tile carries a DNA stem loop pointing downward. (c) The four different tiles are designed to self-assemble into a 2D array displaying parallel lines of Au NPs. (d) Possible ways for the corresponding edge tiles on opposite sides of the 2D array to associate and lead to formation of tubes displaying patterns of Au NPs in stacked rings, single spirals, double spirals, and nested spiral tubes. (e) The different tube conformations were observed in a single TEM image. (Reprinted with permission from [5]. Copyright 2009 American Association for the Advancement of Science.)

tions, with the A tiles forming parallel lines of Au NPs that are all located at the top side of the array, and the C tiles forming parallel lines of stem loops at the bottom side of the array. The designed periodicity between the neighboring A tiles is expected to be ~ 64 nm when the tiles are closely packed side by side. Additionally, the intrinsic curvature of the array is expected to be cancelled because the A and C tiles face one direction, whereas the B and D tiles face the opposite direction. However, in the presence of the 5-nm Au NPs, which have diameters that are comparable to or even greater than the center-to-center distances of the neighboring A tiles within the parallel stripes (4 to 5 nm), the strong electrostatic and steric repulsions between the neighboring Au NPs force the 2D arrays to curl up to avoid direct contact between the particles. This curling will lead to tube formation with the particles displayed on the outer surface of the tubes.

The stem loops on the C tiles in this design were placed on surfaces opposite from the Au NPs as a counterforce to resist tube formation (that is, to increase the energy barrier for bending the 2D array). However, because the Au NPs are much larger than the DNA stem loops, their forces are not perfectly counterbalanced. The tile arrays still have a tendency to curl up into tubes with the stem loops wrapped inside and the Au NPs displayed on the outside. There are a few different ways for the edge tiles to associate in the tube formation [Figure 9.2(d)]. When the edge tiles at one side of the array that associate with the corresponding edge tiles at the opposite side of the array are within the same lines, a tube displaying stacked rings of Au NPs will form. When the corresponding edge tiles that associate are at neighboring lines (with a 1-line offset), tubes displaying a single spiral of nanoparticles will result. Depending on the sign of the offset ($n \rightarrow n + 1$ lines, or $n \rightarrow n - 1$ lines), the spiral can potentially display either a left-handed or right-handed chirality. Similarly, when the corresponding edge tiles that associate are at alternating lines ($n \rightarrow n + 2$ lines, or $n \rightarrow n - 2$ lines), tubes displaying double spirals of nanoparticles will result. When the corresponding edge tiles are at lines with a larger interval ($\Delta n \geq 3$ lines), spiral tubes will be nested. Varieties of tubes with different conformations were observed from the above design [Figure 9.2(e)].

9.2.2 Structural Biomimicry

9.2.2.1 Gecko Foot

Geckos are exceptional in their ability to climb rapidly up smooth vertical surfaces. Microscopy has shown that a gecko's foot has nearly 500,000 keratinous hairs or setae (Figure 9.3) [6]. Each 30- to 130- μm -long seta is only one-tenth the diameter of a human hair and contains hundreds of projections terminating in 0.2- to 0.5- μm spatula-shaped structures. Autumn et al. [7] first directly measured single setal force by using a 2D microelectromechanical systems force

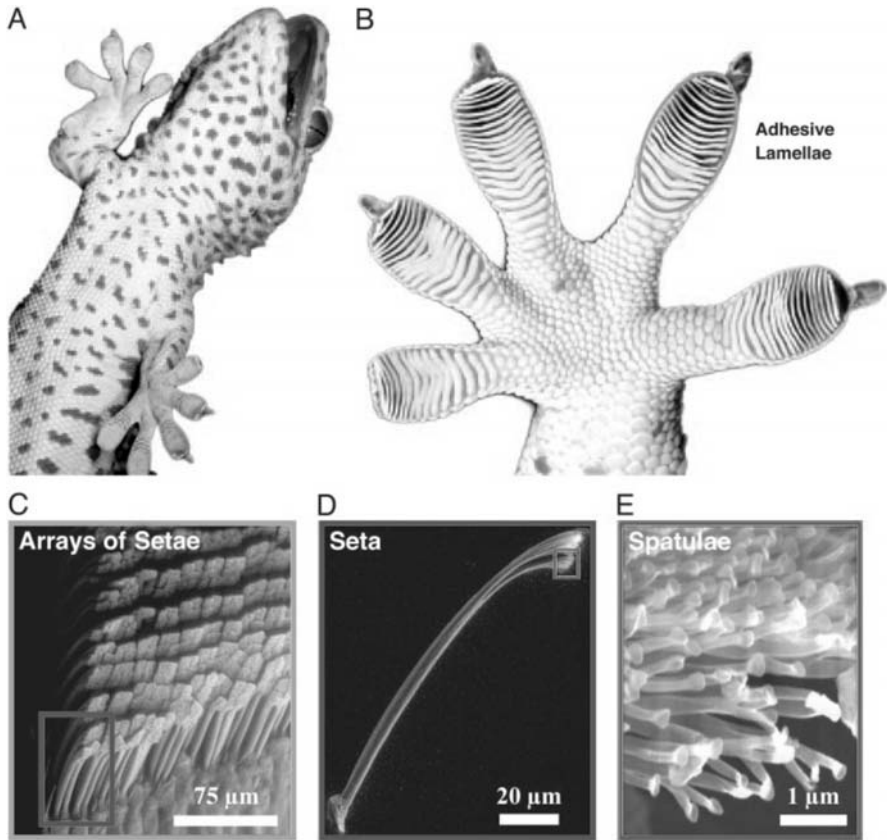


Figure 9.3 Structural hierarchy of the gecko adhesive system. (a) Macrostructure: ventral view of a tokay gecko (*Gekko gekko*) climbing vertical glass. (b) Mesostructure: ventral view of the foot, with adhesive lamellae (scansors) visible as overlapping pads. Note the clean appearance of the adhesive surface. (c) Microstructure: proximal portion of a single lamella, with individual setae visible in an array. (d, e) Nanostructure: single seta with branched structure at upper right, terminating in hundreds of spatular tips. (Reprinted with permission from [6]. Copyright 2005 National Academy of Sciences.)

sensor and a wire as a force gauge. Measurements revealed that a seta is 10 times more effective at adhesion than predicted from maximal estimates on whole animals. Adhesive force values support the hypothesis that individual seta operate by van der Waals forces.

It would be natural to assume that large arrays of tip-like submicrometer objects would mimic a gecko's foot-hair and provide similar adhesive properties. However, one has to take into account that real surfaces are never ideally flat. In fact, it will fail to imitate gecko foot-hair by making large arrays of plastic

tips. The reason is that only a small fraction of tips based on a solid substrate can make physical contact with the opposite surface. Thus, to create a gecko adhesive, one also has to find a way to make hairs sufficiently flexible and to place them on a soft, flexible substrate, so that individual tips can act in unison and attach to uneven surfaces all at the same time [8]. To meet these nontrivial objectives, micropatterned thin polypropylene films may be suitable because they are both robust and flexible and allow a wide range of microfabrication procedures. Figure 9.4(b) shows an example of polypropylene hairs microfabricated using electron-beam lithography [9]. To emphasize the large adhesive force of “gecko tape,” the tape is used to support the weight of a suitably light familiar object [Figure 9.4(a)].

9.2.2.2 Nacre

Rigid biological materials, such as shells, bone, and sponge spicules, have been attractive as models for synthetic structural composites because of their unusual combinations of mechanical properties, such as strength, stiffness, and toughness [10]. Besides their hierarchical structures, another interesting fact is that the smallest building blocks in such materials are generally on the nanometer length scale [11]. Figures 9.5(a, d) show that the enamel of a tooth is made of long, more or less needle-like crystals ≈ 15 to 20 nm thick embedded in soft matrix. Figures 9.5(b, e) show that the nanostructure of bone consists of mineral crystal platelets with a thickness of around a few nanometers embedded in a collagen

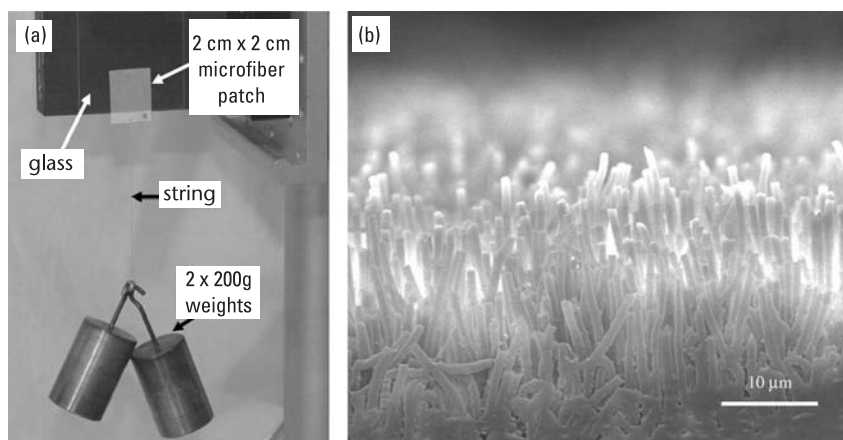


Figure 9.4 (a) A 2-cm^2 microfiber array patch holding two 200g weights (400g) on a vertical, smooth glass slide without normal load. (b) A scanning electron microscope (SEM) image of a microfiber array. The average length and diameter of the microfibers are 18 and $0.6\text{ }\mu\text{m}$, respectively. (Reprinted with permission from [9]. Copyright 2008 Royal Society.)

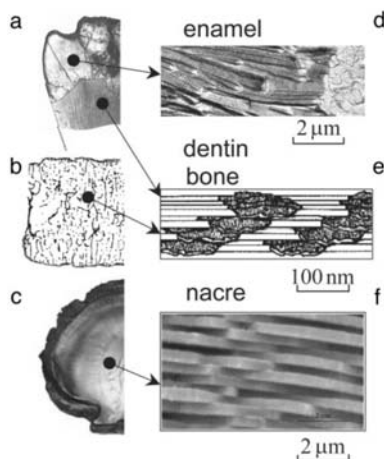


Figure 9.5 Many hard biological tissues, such as (a) tooth, (b) vertebral bone, or (c) shells are made of nanocomposites with hard mineral platelets in a soft (protein) matrix. (d) Enamel is made of long, more or less needle-like crystals ~ 15 to 20 nm thick and $1,000$ nm long, with a relatively small volume fraction of a soft protein matrix. (e) Dentin and bone are made up of plate-like crystals (2 to 4 nm thick, up to 100 nm long) embedded in a collagen-rich protein matrix. The volume ratio of mineral to matrix is on the order of $1:2$. (f) Nacre is made of plate-like crystals (200 to 500 nm thick and a few micrometers long) with a very small amount of soft matrix in between. All of the composites share the structural feature of hard platelets with a very large aspect ratio, arranged in parallel in a brick-and-mortar-like fashion. An unsolved problem is the question of why these crystals are in the nanometer range. (Reprinted with permission from [11]. Copyright 2003 National Academy of Sciences.)

matrix. Figures 9.5(c, f) show the “brick-and-mortar” structure of nacre where the thickness of the aragonite bricks is around a few hundred nanometers.

Ice-templated (IT), porous, layered materials with layers as thin as 1 mm are first fabricated through a freeze-casting process, which involves the controlled unidirectional freezing of ceramic suspensions [12]. These porous scaffolds are then filled with a selected second phase (organic or inorganic) to fabricate dense composites. Then an artificial nacre-like structure is formed. In this method, directional freezing of the ceramic slurries is achieved by pouring them into polytetrafluoroethylene molds placed between two copper cold fingers whose temperature is regulated to control the speed of the solidification front. As in nature, during the freezing of sea water, the ceramic particles concentrate in the space between the ice crystals [Figure 9.6(a)]. When the freezing rate increases, the magnitude of supercooling ahead of the solidifying interface is increased, and as a result the tip radius of the ice crystals decreases. A finer microstructure is thus obtained without affecting the long-range order of the

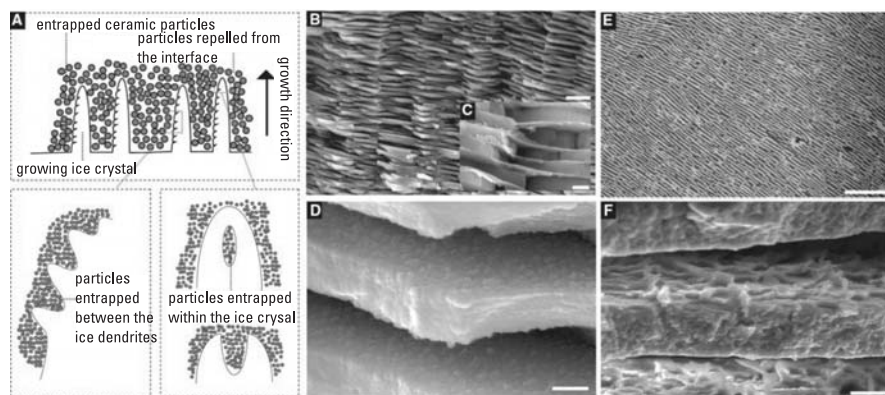


Figure 9.6 Processing principles and materials. While the ceramic slurry is freezing, the growing ice crystals expel the ceramic particles, creating a lamellar microstructure oriented in a direction parallel to the movement of the freezing front (a). For highly concentrated slurries, the interaction between particles becomes critical: A small fraction of particles is entrapped within the ice crystals by tip-splitting and subsequent healing (a), leading to the formation of inorganic bridges between adjacent walls. Dense composites are obtained by infiltrating the porous lamellar ceramic with a second phase (for example, a polymer or a liquid metal). Natural nacre has a bricks-and-mortar-type bridge microstructure where inorganic calcium carbonate layers are held together by organic protein “glue” (b, c); the roughness of the inorganic walls (d) is a key contributor to the final mechanical properties of nacre. The layered microstructure of the IT dense composites resembles that of nacre [for example, the alumina–Al–Si composite in (e)]. The particles entrapped between the ice dendrites generate a characteristic roughness on the lamella surface (f) that mimics that of the inorganic component of nacre. Scale bars: (b) 5 μm , (c) 0.5 μm , (d) 0.3 μm , (e) 300 μm , and (f) 10 μm . (Reprinted with permission from [12]. Copyright 2006 American Association for the Advancement of Science.)

entire structure. Afterwards, the ice is sublimated by freeze drying, such that a ceramic scaffold whose microstructure is a negative replica of the ice is produced [Figure 9.6(f)].

Layer-by-layer (LbL) assembly has also been used to grow artificial nacre [13, 14]. The idea of the technique is very simple. A charged substrate, for instance, negatively charged glass, is immersed in the solution of positively charged polyelectrolyte. After rinsing with water, the latter forms a submonolayer on the surface of the substrate, which switches the surface charge to positive. When it is immersed in the dispersion of negatively charged nanoparticles or other nanocolloid, a new layer is formed, which also switches the surface charge. This makes possible the adsorption of a new layer of polyelectrolyte. The whole cycle can be repeated as many times as desired. Additionally, the chemical nature of the polyelectrolyte and/or the nanocolloid can be varied as the films are being

assembled. This paves the way to preparation of an extremely diverse and flexible family of nanocomposites with a variety of functionalities. For instance, the nanoparticles can be semiconducting, metallic, or magnetic. To add biological functionality to the multilayer, layers of proteins or other biopolymers can be added. Thus, LbL assembly can be considered a very convenient tool for making artificial nacre.

9.2.2.3 Compound Eyes

Compound eyes in nature present intriguing topics in physiological optics because of their unique optical scheme for imaging [15]. For example, a bee's eye has thousands of integrated optical units called ommatidia spherically arranged along a curvilinear surface so that each unit points in a different direction [Figure 9.7(a)]. Each ommatidium consists of a light-diffracting facet lens, a crystalline cone, and photoreceptor cells with a wave-guiding rhabdom [Figure 9.7(b)]. The omnidirectionally arranged ommatidium collects incident light with a narrow range of angular acceptance and independently contributes to the capability of wide field-of-view (FOV) detection.

In [15], biologically inspired artificial compound eyes were developed in a small form factor with 3D configurations. These biomimetic compound eyes are anatomically as well as functionally close to natural compound eyes [Figure 9.7(c)]. The artificial ommatidium consists of a honeycomb-packed hexagonal microlens with a low Fresnel number ($N_F < 10$), a cuvette-shaped polymer cone, and a polymer waveguide that has a higher index solid core surrounded by a lower index solid cladding in the polymer resin [Figure 9.7(d)]. Three-dimensional polymer synthesis of an artificial compound eye can be realized through microlens templating, reconfigurable microtemplating, and self-writing in a photosensitive polymer resin. Each ommatidium was omnidirectionally arranged in a hemispherical polymer dome. Like the crystalline cone in nature, the polymer cone helps guide the focused light into the polymer waveguide, and subsequently the guided light arrives at the end of the waveguide core. Lastly, light detection can be done by photodetector arrays. In 3D implementation, microlens-assisted self-writing and polymer replication processes were used to minimize the lens-waveguide coupling loss and to realize a spherical configuration, respectively.

In the aforementioned approach, nanostructures were first imprinted atop a planar and elastic membrane; then, the medium was deformed by a negative pressure in a sealed environment. Compatibility with a variety of lithographic processes and benefits of good accuracy in alignment imply great advantages. Design complexity, however, might be associated with their process when a freestanding object is demanded or the target object has a tubular or irregular shape or geometry. A key gap exists in the development of an embossing/imprinting-based patterning technique for freestanding soft objects with various

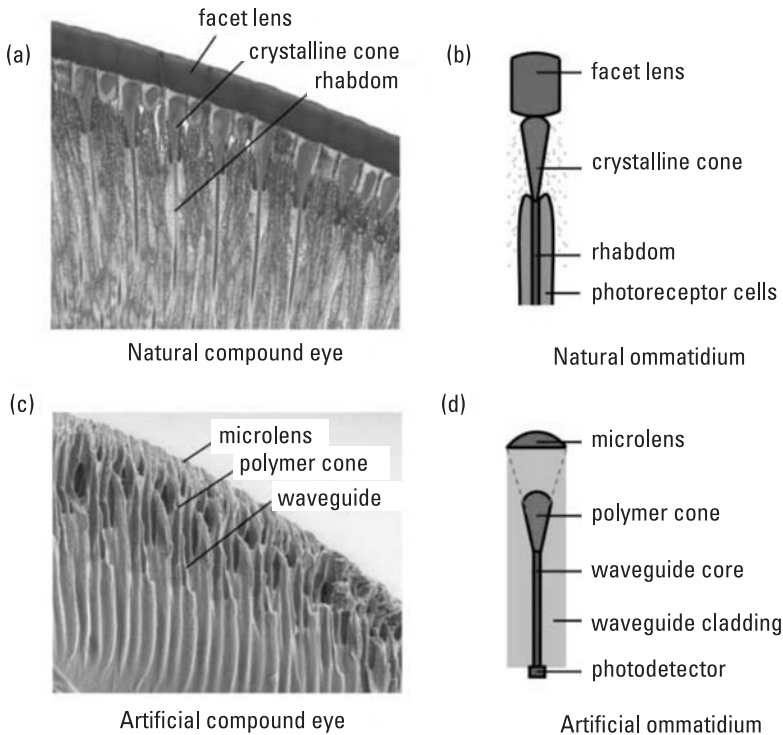


Figure 9.7 Anatomical comparisons between a natural compound eye and an artificial compound eye described from the cross sections. (a) An optical micrograph of a honeybee's apposition compound eye. As an individual optical unit, (b) a natural ommatidium consists of a facet lens, a crystalline cone, and photoreceptor cells with a wave-guiding rhabdom. (c) A scanning electron micrograph of an artificial compound eye and (d) an artificial ommatidium consisting of a microlens, a polymer cone, and an optical waveguide that has a higher index core surrounded by a lower index cladding in a polymer resin. Light impinging onto a microlens is coupled with polymer cones and waveguides and then guided to the end of the waveguide. (Reprinted with permission from [15]. Copyright 2006 American Association for the Advancement of Science.)

curvatures, such as tubular 1D, spherical 2D, or more complex structures, that can, at the same time, deliver good accuracy in alignment steps and be compatible with conventional planar lithography platforms. Lima et al. [16] developed a new technique to create micro- and nanometer scale structures on curved freestanding objects by combining embossing/imprinting lithography approaches with mechanical loadings on elastic films. The process of embossing and imprinting generates small structures, and mechanical loading determines the shape or geometry of the final object. As a result, a portion of the tubes

with radii between 0.5 and 3.5 mm and a portion of the spheres with radii between 2.4 and 7.0 mm were fabricated with grating line features (period of 700 nm) and microlens array features (lens radius of $2.5\ \mu\text{m}$) atop, respectively.

Approximations to portions of tubes were fabricated following the process flow shown in Figure 9.8(a). In this process, patterned PDMS thin film was first mounted on a 1D stretcher [Figure 9.8(c)], with nanostructures facing down [step 1, Figure 9.8(a)]. The nanostructures were fabricated following conventional embossing/imprinting lithography, after which grating lines with a pitch of 700 nm and a linewidth of 350 nm were obtained. The stretcher is then allowed to deform the film until reaching a desired strain; then a fresh PDMS precursor is cast atop the stretched film and baked to form a bilayer membrane structure [step 2, Figure 9.8(a)]. During the baking process, the PDMS precursor penetrates into the stretched solid film, which gives rise to a tight bond between the two layers. After the stretcher is released, the strain in the bottom PDMS layer is relaxed, which compresses the top PDMS layer and generates a freestanding and soft portion of tube [step 3, Figure 9.8(a)]. Finally, epoxy is used to fill the soft portion of the tube, which revealed to us a portion of a rigid cylinder, with surface nanostructures copied from the inner side of the soft portion of the tube [step 4, Figure 9.8(a)].

Figure 9.9 presents both optical and SEM micrographs of the final freestanding portion of the tube [insets on left, Figures 9.9(a, b)], rigid replica [insets on right, Figures 9.9(a, b)], and surface nanostructures [Figures 9.9(a, b)]. These two rigid replicas have radii of 0.5 and 3.5 mm, respectively. The SEM images in Figure 9.9 show the nanostructures on the surfaces of rigid replicas. These tiny structures are negative copies of the gratings inside the freestanding object, with a linewidth of 350 nm and a pitch dimension of 700 nm. It is worthwhile to note that the variations seen in the image's contrast from top to bottom in Figure 9.9(b) indicate the inward bending of the rigid replica with respect to the paper.

9.3 Nanostructures in Biodiagnostics

In the last 10 years, the field of molecular diagnostics has witnessed an explosion of interest in the use of nanomaterials in assays for gases, metal ions, and DNA and protein markers for many diseases [17]. Intense research has been fueled by the need for practical, robust, and highly sensitive and selective detection agents that can address the deficiencies of conventional technologies. Chemists are playing an important role in designing and fabricating new materials for application in diagnostic assays. In certain cases assays based on nanomaterials have offered significant advantages over conventional diagnostic systems with regard to assay sensitivity, selectivity, and practicality.

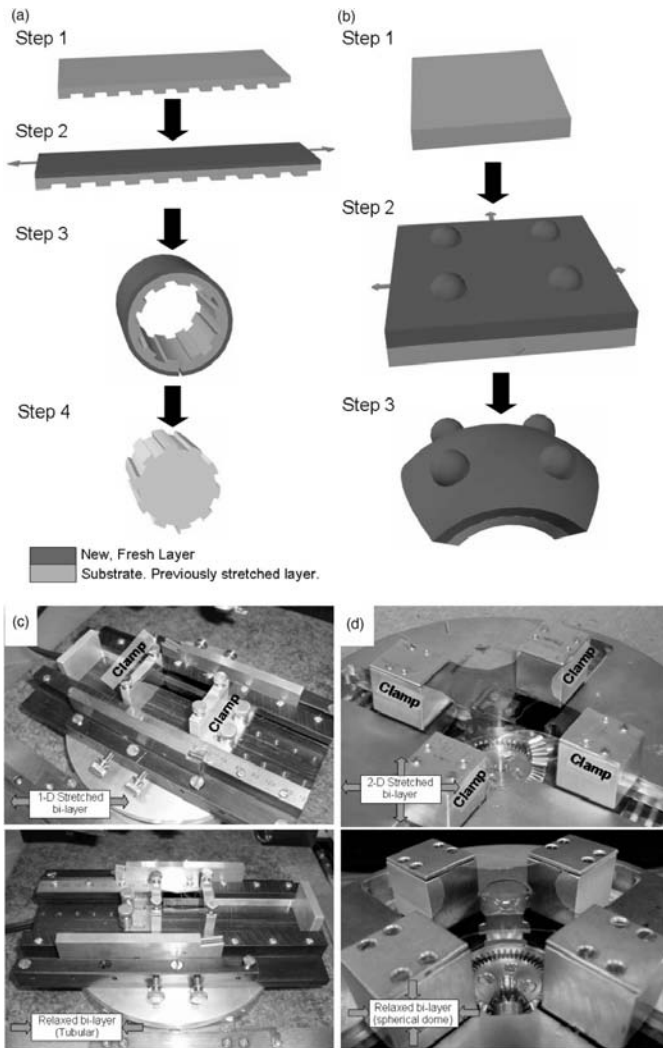


Figure 9.8 (a) Schematic process flow to produce nanostructures on cylindrical objects: Step 1: embossing/imprinting patterned soft film with features on the bottom; step 2: stretch the film in 1D fashion and cast a liquid elastomer precursor, then bake to form a bilayer structure; step 3: relax the bilayer to form a portion of tube; and step 4: replicate the freestanding portion of the tube into a portion of a rigid cylinder. (b) Process flow to produce nanostructures on spherical objects: Step 1: prepare a soft film as substrate; step 2: stretch the film in 2D fashion and emboss/imprint a new layer with patterns; and step 3: relax the bilayer to form a portion of the sphere. Homemade stretching tools are used for (c) 1D and (d) 2D manipulation. Elastic film can be fixed between solid bars, as indicated in (c) and (d). Translation of these bars introduces 1D or 2D deformation to the soft films in between. (Reprinted with permission from [16]. Copyright 2007 American Institute of Physics.)

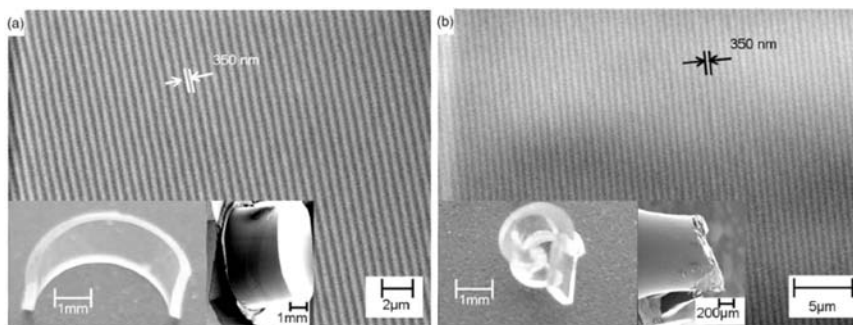


Figure 9.9 SEM images of grating line features fabricated atop the curved surface of portions of rigid cylinders. Insets: freestanding bilayer portions of tubes (a) and rigid replica (b). The freestanding tubular structures were received by applying uniaxial strain at 40% [(a) inset] and 80% [(b) inset], respectively. (Reprinted with permission from [16]. Copyright 2007 American Institute of Physics.)

9.3.1 Nanoparticle-Based Detection Methods

An early indication of the potential of nanomaterials as biodetection agents, beyond conventional histochemical staining, was reported in 1996 [18] with the observation that oligonucleotide-modified nanoparticles and sequence-specific particle assembly events, induced by target DNA, could be used to generate materials with unusual optical and melting properties. Specifically, when 13-nm gold particles were used in the assay, the color of the solution changed from red to blue upon the analyte-directed aggregation of gold nanoparticles, a consequence of interacting particle surface plasmons and aggregate scattering properties. This simple phenomenon pointed toward the use of nanoparticles as DNA detection agents in a type of “litmus test” for nucleic acid targets, and indeed, it was found that spotting the solution onto a white support enhanced the colorimetric change and provided a permanent record for each test (Figure 9.10) [17, 19].

Further studies indicated that the melting profiles of the nanoparticle-labeled DNA aggregates were extraordinarily sharp, occurring over a temperature range much more narrow than the transition for unlabeled or conventional fluorophore-labeled DNA (Figure 9.10). These two observations, both consequences of the high surface area and unique optical activity of the gold nanoparticles, created worldwide interest in exploring the potential for designer nanomaterials in biodiagnostic applications. The colorimetric change pointed to a simple and inexpensive way of diagnosing disease, and the unanticipated sharp melting profile suggested that assays based on such nanostructures should have higher selectivities than the conventional molecular fluorophore-labeled structures that exhibit broad melting profiles when hybridized with complemen-

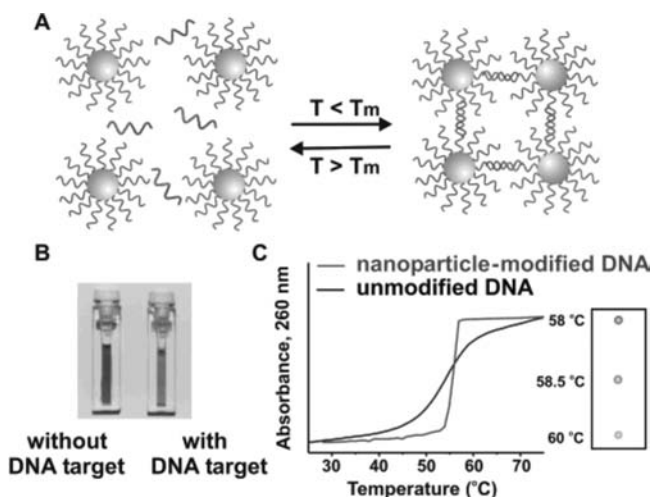


Figure 9.10 (a) In the presence of complementary target DNA, oligonucleotide-functionalized gold nanoparticles will aggregate, resulting (b) in a change of solution color from red to blue. (c) The aggregation process can be monitored using UV-vis spectroscopy or simply by spotting the solution on a silica support. (Reprinted with permission from [17]. Copyright 2005 American Chemical Society.)

tary DNA. The sharp melting transitions associated with these nanoparticle probes derive from the dense loading of oligonucleotides on their surfaces and their ability to bind to complementary DNA in a highly cooperative manner. These properties have not been observed with microparticle probes, partly because the loading efficiency of oligonucleotides does not compare with the gold nanoparticle-thiol system. It is worth noting that colorimetric responses have been utilized in viral detection systems based on supramolecular polydiacetylene liposomes.

9.3.2 Nanowire- and Nanotube-Based Detection Methods

Nanotubes and nanowires are being explored as new signal transduction motifs in the electrical detection of DNA as they have been for the detection of gases, small molecules, and proteins (*vide infra*). For example, Lieber and colleagues [20] demonstrated that silicon nanowires functionalized with peptide nucleic acid (PNA) can be used for real-time, label-free detection of DNA. In their assay the conductance of a PNA-functionalized silicon nanowire bridging two electrodes is measured in the presence of target DNA and mutant DNA with three consecutive base deletions. Introduction of target DNA into the assay caused a rapid and immediate change in conductance, whereas the effect of mutant DNA was negligible. Furthermore, the conductance changes scale with

target concentration, and target DNA can be detected at concentrations as low as 10 fM. In the case of nanotubes, Lieber and colleagues [21] showed that specific sequences of kilobase-size DNA can be detected using single-wall carbon nanotube (SWCNT) atomic force microscopy (AFM) probes. Specifically, they marked particular sequences along the DNA strand with streptavidin-labeled complementary DNA probes and then used AFM to identify the streptavidin and thus the location of the target sequences. This technique enabled the detection of specific haplotypes that code for genetic disorders.

9.4 Nanostructures in Cells Study

The cell is the core of biology: It is the smallest unit that is alive. In their normal environment, cells are subject to multiple cues that vary in time and space, including gradients of cytokines and secreted proteins from neighboring cells, biochemical and mechanical interactions with the extracellular matrix (ECM), and direct cell–cell contacts [22]. Nanofabricated systems can present cells with these cues in a controllable and reproducible fashion that cannot easily be achieved by standard tissue culture, and can be used to link a cell culture with integrated analytical devices that can probe the biochemical processes that govern cell behavior.

9.4.1 Microarray Platform as a Research Tool

The force exerted on the ECM by cells can be measured in several ways. A particularly powerful method involves measuring the deflection of arrays of micrometer-sized vertical elastomer posts. To examine the adhesion of cells on the mPADs, Tan et al. [23] uniformly adsorbed fibronectin and seeded cells onto a substrate containing posts 3 μm in diameter, 11 μm in height, with 6- μm spacing. Cells attached, spread across multiple posts, and deflected underlying posts [Figure 9.11(a)]. However, cells spread down the length of the posts and, therefore, were exposed to a range of local compliances. The deflection of the posts therefore does not quantitatively reflect the applied traction force. To restrict cell adhesion to the tips of the posts, and thereby precisely define the surface mechanics, we used microcontact printing to deliver fibronectin from a stamp onto the tips of the posts [Figure 9.11(b)], and adsorbed Pluronic F127 onto the remaining unstamped regions of the array to block nonspecific protein adsorption and cell adhesion. Using this method, we can print fibronectin onto specific posts within the array to spatially pattern the adhesiveness of the surface toward cells [Figure 9.11(c)]. On substrates where the tips of all of the posts were printed with fibronectin by using a flat stamp, cells attached, spread, and migrated selectively across the tips of the posts, bending the posts centripe-

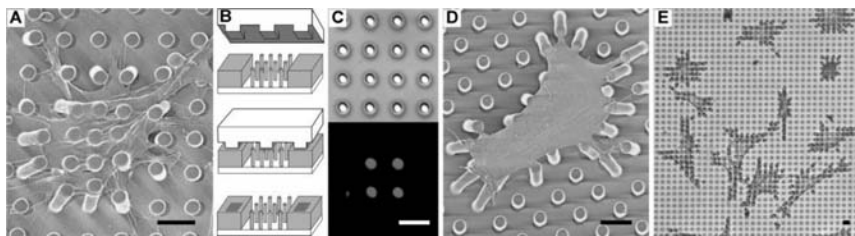


Figure 9.11 Cell culture on arrays of posts. (a) Scanning electron micrograph of a representative smooth muscle cell attached to an array of posts that was uniformly coated with fibronectin. Cells attached at multiple points along the posts as well as the base of the substrates. (b) Schematic of microcontact printing of protein, pre-coated on a PDMS stamp, onto the tips of the posts. (c) Differential interference contrast (upper) and immunofluorescence (lower) micrographs of the same region of posts where a 2×2 array of posts has been printed with fibronectin. (d) Scanning electron micrograph and (e) phase-contrast micrograph of representative smooth muscle cells attached to posts where only the tips of the posts have been printed with fibronectin by using a flat PDMS stamp. Cells deflected posts maximally during the 1h to 2h period after plating, were fully spread after 2h, and were fixed and critical point dried 4h after plating. (Scale bars indicate $10 \mu\text{m}$.) (Reprinted with permission from [23]. Copyright 2003 National Academy of Sciences.)

tally toward the interior of the cell [Figure 9.11(d)]. The general morphology of the cells on posts was similar to that of cells cultured on planar substrates [Figure 9.11(e)].

The microarrays of biological probes are produced via sequential illumination of a photosensitive reagent through differently addressed masks. Alternative lithography techniques, such as contact lithography and ink-jet printing, have produced hierarchical microstructures on a broad range of materials including nonflat surfaces. Further miniaturization using these methods, however, is not trivial due to the precision requirements in mask making and alignments between steps. Scanning probe lithography can provide high spatial precision in fabrication and characterization of nanostructures; however, this method suffers the drawback of low throughput. Inspired by the heat-induced miniaturization protocols for polymer materials and the surface-responsive nature of elastomers, Ouyang et al. [24] demonstrated a new platform for miniaturization. This platform, dubbed *stepwise contraction and adsorption nanolithography* (SCAN), should in principle be able to extend the advantages of multicomponent micro-fabrication into smaller dimensions, as well as to reach the spatial precision afforded by scanning probe lithography.

The miniaturization principle of SCAN is illustrated in Figure 9.12(a) using an analogy with a balloon, and the key steps are shown schematically in Figure 9.12(b). A synthetic polymer with high elasticity is used as the stamp,

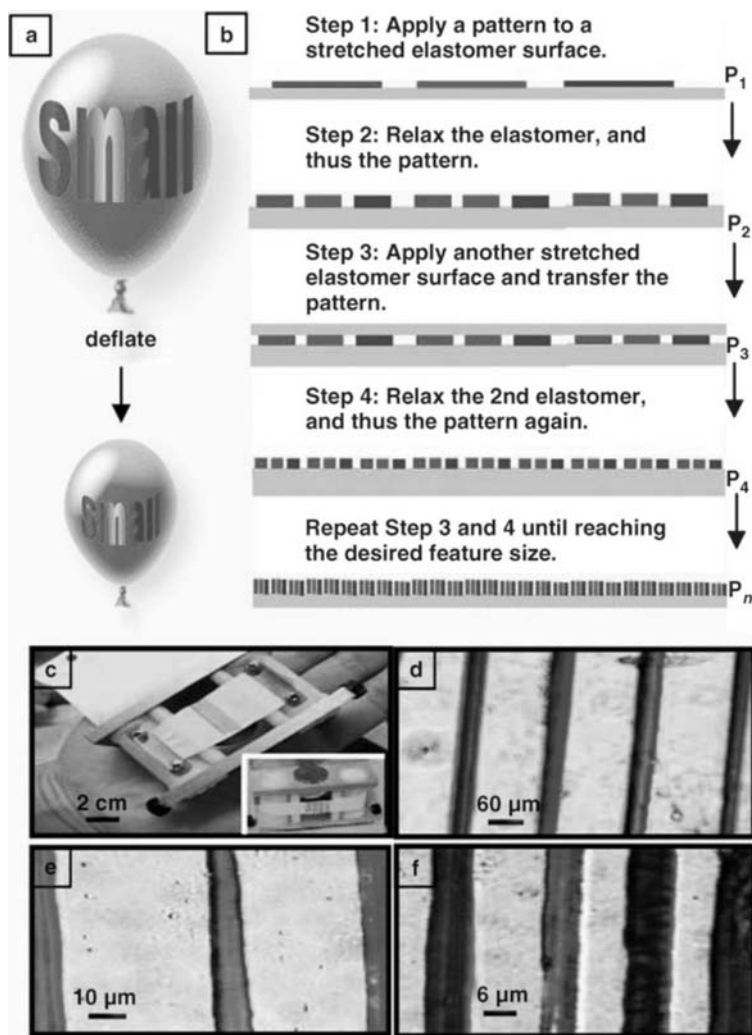


Figure 9.12 The concept of SCAN: (a) Balloon analogy of SCAN, in which the word “Small” on a balloon is miniaturized upon deflation. (b) Schematic of the procedure. (c) A photograph of the 1D SCAN stretcher with original ink lines drawn on a latex surface. (d) Optical image of the lines after one SCAN cycle. (e, f) Images captured after the second and third SCAN cycles, respectively. (Reprinted with permission from [24]. Copyright 2006, Wiley Interscience.)

and the miniaturization process is derived from the uniform shrinkage of the central area of this elongated substrate. The fabrication procedure includes (1) the design and production of an original microstructure (P_1) on a stretched rubber surface; (2) the miniaturization of the pattern using the following cycle:

relaxing the elongated rubber to form a smaller pattern (P_2) and transferring the new structure to another stretched elastomer through contact; (3) repeating this cycle until the final pattern size (P_n) is obtained; and (4) the final transferring of the pattern (P_n) onto the desired surfaces. SCAN can be performed in one or two dimensions depending on the geometry requirements of the final products. For example, a 1D SCAN process would be selected for generating arrays of lines [Figure 9.12(c–f)].

9.5 Tissue Engineering and Drug Delivery

Micro- and nanoscale technologies are emerging as powerful enabling tools for tissue engineering and drug delivery. In tissue engineering, micro- and nanotechnologies can be used to fabricate biomimetic scaffolds with increased complexity and vascularization. Furthermore, these technologies can be used to control the cellular microenvironment (i.e., cell–cell, cell–matrix, and cell–soluble factor interactions) in a reproducible manner and with high temporal and spatial resolution. In drug delivery, miniaturized platforms based on micro- and nanotechnology can be used to precisely control the fluid flow, enable high-throughput screening, and minimize sample or reagent volumes [25]. In addition, these systems enhance reproducibility and significantly reduce reaction times.

Nanotechnology can be used to create nanofibers, nanopatterns, and controlled-release nanoparticles with applications in tissue engineering. These techniques are particularly useful for mimicking native tissues because many biological structures, such as ECM fibers, are in the range of tens of nanometers [26]. For example, polymeric nanofibers that mimic collagen fibers can be fabricated by electrospinning and self-assembly. In general, the synthesis of nanostructured materials can be generated by using one of two approaches. In one approach, nanomaterials are synthesized by miniaturizing existing materials with nanoscale resolution [26]. These techniques include nanopatterning and electrospinning. Electrospun nanofibers are versatile tools to fabricate tissue engineering scaffolds with biomimetic mechanical, chemical, and biological properties [Figures 9.13(a–c)]. In the other approach, molecular buildup, such as self-assembly and LbL deposition, can be used to generate nanomaterials [Figures 9.13(d–f)] [27, 28].

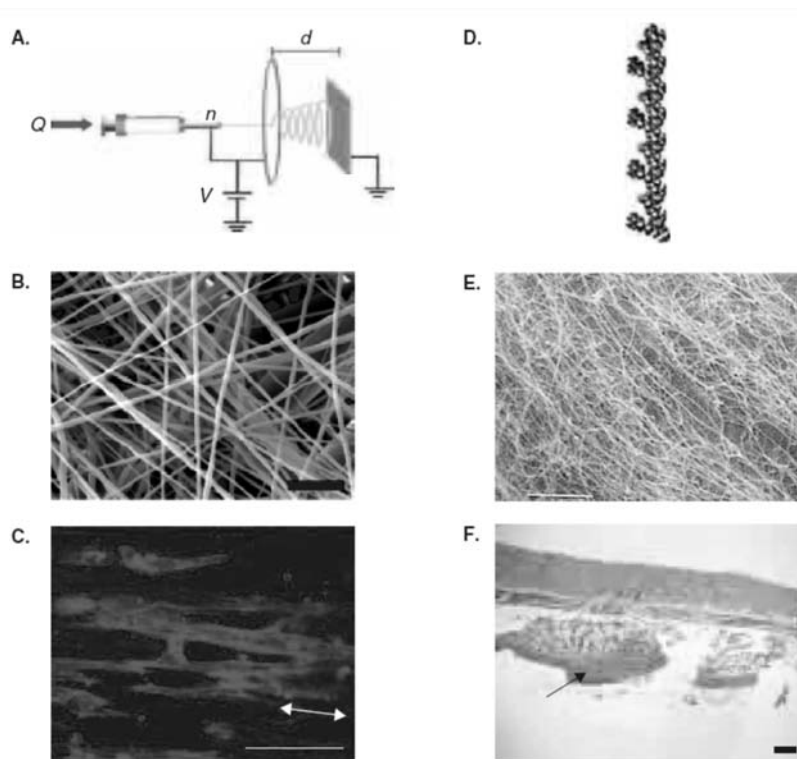


Figure 9.13 Polymeric nanofibers for tissue-engineering applications. To synthesize polymeric nanofibers, electrospinning and self-assembly can be used. (a) Scheme of a typical electrospinning setup. (b) SEM images of electrospun nanofibers. Scale bar is $10\ \mu\text{m}$. (c) A confocal laser scanning microscope image of cardiac myocytes on predefined oriented fibers of PLGA + PEG-poly(D, L-lactide) (PLA) diblock copolymer. Scale bar is $20\ \mu\text{m}$. (d) Schematic image of a peptide that can self-assemble to form 3D hydrogels. (e) SEM image of self-assembled PA nanofiber networks containing BMP-2. (f) 3D nanofiber scaffolds with BMP-2 significantly induced ectopic bone formation around the injected site. Arrows indicate the newly generated ectopic bone. Scale bar is $1\ \text{mm}$. BMP-2: bone morphogenic protein-2; PA: peptide amphiphile; PEG: poly(ethylene glycol); PLA: poly(lactic acid); PLGA: poly(lactic-co-glycolide). (Reprinted with permission from [28]. Copyright 2007 National Academy of Sciences.)

References

- [1] Whitesides, G. M. "The 'Right' Size in Nanobiotechnology," *Nature Biotechnology*, Vol. 21, 2003, pp. 1161–1165.
- [2] Seeman, N. C. "DNA Nicks and Nodes and Nanotechnology," *Nano Letters*, Vol. 1, 2001, pp. 20–26.

-
- [3] Seeman, N. C. "DNA in a Material World," *Nature*, Vol. 421, 2003, pp. 427–431.
 - [4] Aldaye, F. A., A. L. Palmer, and H. F. Sleiman, "Assembling Materials with DNA as the Guide," *Science*, Vol. 321, 2008, pp. 1795–1799.
 - [5] Sharma, J., et al. "Control of Self-Assembly of DNA Tubules Through Integration of Gold Nanoparticles," *Science*, Vol. 323, 2009, pp. 112–116.
 - [6] Hansen, W. R., and K. Autumn, "Evidence for Self-Cleaning in Gecko Setae," *Proc. of the National Academy of Sciences of the USA*, Vol. 102, 2005, pp. 385–389.
 - [7] Autumn, K., et al., "Adhesive Force of a Single Gecko Foot-Hair," *Nature*, Vol. 405, 2000, pp. 681–685.
 - [8] Geim, A. K., et al. "Microfabricated Adhesive Mimicking Gecko Foot-Hair," *Nature Material*, Vol. 2, 2003, pp. 461–463.
 - [9] Lee, J., et al., "Sliding Induced Adhesion of Stiff Polymer Microfiber Arrays: 1. Macroscale Behaviour," *J. of the Royal Society Interface*, Vol. 5, 2008, pp. 835–844.
 - [10] Mayer, G., "Rigid Biological Systems as Models for Synthetic Composites," *Science*, Vol. 310, 2005, pp. 1144–1147.
 - [11] Gao, H. J., et al. "Materials Become Insensitive to Flaws at Nanoscale: Lessons from Nature," *Proc. of the National Academy of Sciences of the USA*, Vol. 100, 2003, pp. 5597–5600.
 - [12] Deville, S., et al., "Freezing as a Path to Build Complex Composites," *Science*, Vol. 311, 2006, pp. 515–518.
 - [13] Tang, Z. Y., et al., "Nanostructured Artificial Nacre," *Nature Materials*, Vol. 2, 2003, pp. 413–418.
 - [14] Podsiadlo, P., et al., "Ultrastrong and Stiff Layered Polymer Nanocomposites," *Science*, Vol. 318, 2007, pp. 80–83.
 - [15] Jeong, K. H., J. Kim, and L. P. Lee, "Biologically Inspired Artificial Compound Eyes," *Science*, Vol. 312, 2006, pp. 557–561.
 - [16] Lima, O., et al., "Creating Micro- and Nanostructures on Tubular and Spherical Surfaces," *J. of Vacuum Science & Technology B*, Vol. 25, 2007, pp. 2412–2418.
 - [17] Rosi, N. L., and C. A. Mirkin, "Nanostructures in Biodiagnostics," *Chemical Reviews*, Vol. 105, 2005, pp. 1547–1562.
 - [18] Mirkin, C. A., et al., "A DNA-Based Method for Rationally Assembling Nanoparticles into Macroscopic Materials," *Nature*, Vol. 382, 1996, pp. 607–609.
 - [19] Elghanian, R., et al., "Selective Colorimetric Detection of Polynucleotides Based on the Distance-Dependent Optical Properties of Gold Nanoparticles," *Science*, Vol. 277, 1997, pp. 1078–1081.
 - [20] Hahm, J.-I., and C. M. Lieber, "Direct Ultrasensitive Electrical Detection of DNA and DNA Sequence Variations Using Nanowire Nanosensors," *Nano Letters*, Vol. 4, 2004, pp. 51–54.
 - [21] Woolley, A. T., et al., "Direct Haplotyping of Kilobase-Size DNA Using Carbon Nanotube Probes," *Nature Biotechnology*, Vol. 18, 2000, pp. 760–763.

- [22] Ei-Ali, J., P. K. Sorger, and K. F. Jensen, "Cells on Chips," *Nature*, Vol. 442, 2006, pp. 403–411.
- [23] Tan, J. L., et al., "Cells Lying on a Bed of Microneedles: An Approach," *Proc. of the National Academy of Sciences of the USA*, Vol. 100, 2003, pp. 1484–1489.
- [24] Ouyang, Z. Q., et al., "A Simple Miniaturization Protocol to Produce Multicomponent Micro- and Nanostructures," *Small*, Vol. 2, 2006, pp. 884–887.
- [25] Santini, J. T., Jr., M. J. Cima, and R. Langer, "A Controlled-Release Microchip," *Nature*, Vol. 397, 1999, pp. 335–338.
- [26] Pham, Q. P., U. Sharma, and A. G. Mikos, "Electrospinning of Polymeric Nanofibers for Tissue Engineering Applications: A Review," *Tissue Engineering*, Vol. 12, 2006, pp. 1197–1211.
- [27] Zhang, S., "Fabrication of Novel Biomaterials Through Molecular Self-Assembly," *Nature Biotechnology*, Vol. 21, 2003, pp. 1171–1178.
- [28] Chung, B. G., L. F. Kang, and A. Khademhosseini, "Micro- and Nanoscale Technologies for Tissue Engineering and Drug Discovery Applications," *Expert Opinion on Drug Discovery*, Vol. 2, 2007, pp. 1–16.

10

Nanostructural Materials

With the emergence of advanced materials processing and characterization techniques, nanoscale research has become a reality. At the nanoscale or even atomic level, not only do the electronic properties of materials exhibit size effects due to quantum mechanics, but the mechanical properties of materials also move from continuum based to discrete, leading to many unique mechanical properties, some of which even approach theoretical limits of strength. This also leads to some special applications for nanostructural materials.

In this chapter we discuss nanostructural materials from several aspects. First, we introduce nanograin-sized structural materials—traditional metals and alloys that are grain refined to nanometer size. Second, we discuss nanoindentation—one of the most used methods to study mechanical behavior on the nanometer scale. Third, we study the mechanical behavior and instability of nanostructures.

10.1 Nanograin-Sized Structural Materials

Grain refinement of structural materials has long been a topic of study in the metallurgy field. With today's advanced thermal and thermal mechanical processing techniques, we are finally able to refine grain size down to the sub-100-nm region, which we can officially call nanograin-sized structural materials. In this section, we first examine why we need grain refinement and then discuss how to refine grain size to the sub-100-nm region.

10.1.1 Why Grain Refinement?

It has long been known that a material's strength and toughness have an inverse relationship with each other, meaning that an increase in strength will inevitably lead to a decrease in toughness. On the other hand, grain refinement is the only method that can contribute to both spontaneously. The yield strength of a material commonly obeys the relation known as the Hall-Petch equation [1, 2]:

$$\sigma_y = \sigma_i + k_y d^{-1/2} \quad (10.1)$$

where σ_y is the yield strength, σ_i is the internal friction, k_y is the Hall-Petch slope representing the resistance against the slip propagation across a grain boundary, and d is the grain size. The apparent fracture strength σ_f is derived from dislocation theory [3, 4] combined with classical Griffith theory [5] for a brittle crack initiated around a dislocation pile-up at the grain boundary with the following result:

$$\sigma_f = 4G\gamma_m k_y^{-1} d^{-1/2} \quad (10.2)$$

where σ_f is the apparent fracture strength, G is the shear modulus, γ_m is the surface energy, k_y is the Hall-Petch slope, and d is the grain size. Because the internal friction σ_i is a strong function of temperature, the yield strength σ_y of a ferric alloy increases significantly with decreasing temperature. On the other hand, the apparent fracture strength σ_f only changes slightly with temperature, resulting in a typical ductile–brittle transition in ferritic alloys at temperature T_C where $\sigma_y = \sigma_f$ as follows:

$$T_C = C^{-1} \ln \left[\frac{Bk_y d^{1/2}}{(\beta G\gamma_m - k_y^{-2})} \right] \quad (10.3)$$

where C , B , and β are constants [6]. Therefore, the transition temperature can be depressed by reducing the Hall-Petch slope k_y , increasing the surface energy γ_m , and refining the grain size. However, from (10.1) and (10.2), it is apparent that grain refinement is the only means by which both yield strength and apparent fracture strength can be increased and the transition temperature can be simultaneously lowered.

Next we discuss grain refinement on the most common structural alloy—steel.

10.1.2 General Approaches for Grain Refinement

As discussed in the preceding section, strength and toughness characterize a material's intrinsic resistance to plastic deformation and cleavage fracture, respec-

tively. Before we discuss grain refinement to improve strength and toughness, we need to define effective grain size in term of transgranular cleavage fracture and plastic deformation.

It is well known that ferritic steel cleaves along the $\{100\}$ planes. Therefore, the effective grain size, in terms of transgranular cleavage fracture, can be defined as the coherent length on crystallographic cleavage planes, or the mean free path of crack propagation along the $\{100\}$ planes in BCC ferritic steel. On the other hand, plastic deformation usually occurs when dislocations glide along the $\{110\}$ planes. Therefore, the effective grain size for plastic transformation is the coherent length of the $\{110\}$ planes. It has been found that these two effective grain sizes of ferritic steel may not be the same and thus may not respond the same way to materials processing for grain refinement.

For example, in lath martensitic steels, the effective grain size is rather complicated and is, in fact, a function of prior thermal processes. A typical microstructure of lath martensitic steel is shown in Figure 10.1 [7]. The prior

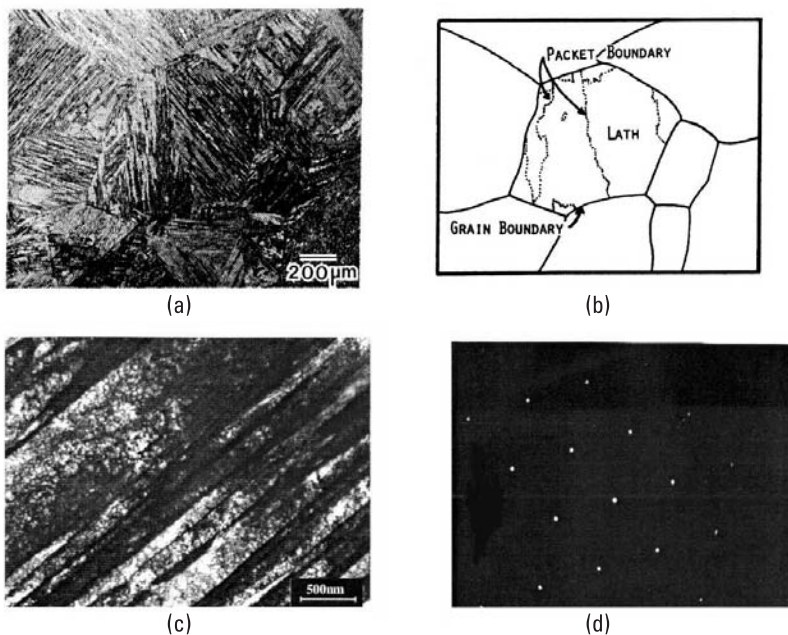


Figure 10.1 The microstructure of a lath martensitic steel: (a) optical microstructure of lath martensitic steel; (b) illustration of subdivided packet within a prior austenitic grain; (c) dislocated lath martensite within a packet showing alignment of laths of less than 0.5 mm thickness; and (d) diffraction pattern showing that the laths are almost identical in crystallographic orientation. (From J. W. Morris, Z. Guo, C. Krenn, Y. H. Kim, *ISIJ International*, Vol. 41, No. 6, 2001.)

austenitic grains are subdivided into packets of thin, closely aligned martensitic laths. The structure within the packet appears to be very fine since the laths are at the submicron or even 100-nm level in cross section. However, the diffraction pattern obtained from the interior of a packet [Figure 10.1(d)] revealed that the laths within a packet are in close crystallographic alignment and the lath boundaries are only low-angle boundaries. Hence, the packet is essentially a single crystal of martensite, and its size will define the effective grain size. This is illustrated in Figure 10.2, which is a profile micrograph of cleavage fracture in an as-quenched 6Ni steel that was treated to produce very large packets and broken at 77K [7]. The cleavage facets follow the $\{100\}$ planes that are common to the laths in a packet, diverging at packet boundaries.

Generally there are two different ways to accomplish grain refinement: thermal mechanical processing or thermal processing. Thermomechanical processing is a combination of mechanical work to obtain large amount of plastic deformation and thermal exposure to induce recrystallization. This method has received dramatic emphasis in recent years with the launch of large “supersteel” projects in Japan, Korea, and China that are intended to develop the next generation of structural steels [8, 9]. It has been shown that exceptional strength and toughness combinations can be achieved in low-alloy steels by refining the

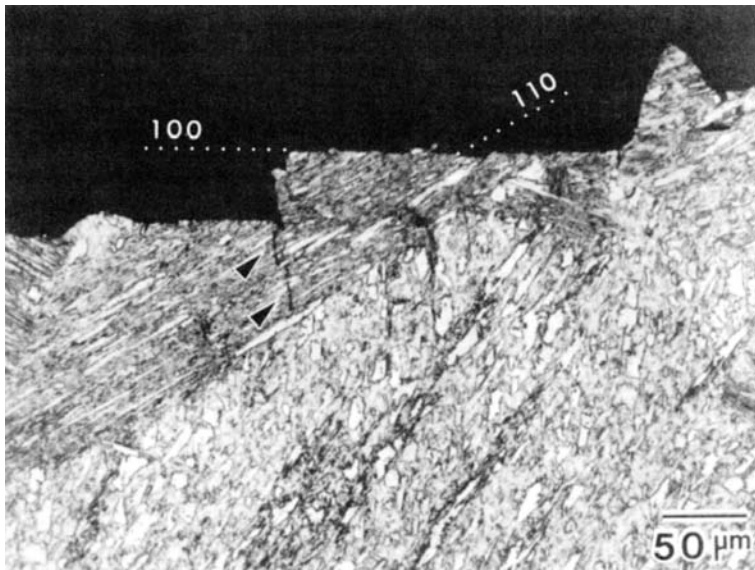


Figure 10.2 Scanning electron micrograph of etched cross section of lath martensitic steel broken below T_B . The cleavage cracks follow common $\{100\}$ planes across parallel laths in a packet, indicating that the effective grain size is the packet size. (Reprinted with permission from [10]. Copyright *ISI/J*, 2001.)

grain size down to the sub micron or even 100-nm range. However, there are still some barriers to implementing these results successfully. It is very difficult to achieve the large amount of uniform plastic deformation that is necessary to produce ultrafine grain size through the thickness of plate steel. This is particularly true when the product of interest is a high-strength, high-alloy thick plate.

The second approach to grain refinement, the thermal method, is particularly applicable to high-strength thick plates, bars, and weldments. These are cyclic thermal treatments that exploit the properties of martensitic transformation and produce ultrafine grain size in lath martensitic steel [11–13]. For example, during the diffusionless martensitic transformation, ferritic steel always exhibits a coherent crystallographic relationship with the parent phase. However, under certain thermal process conditions, it has been revealed crystallographically that the cleavage planes $\{001\}$ exhibit large angle misorientation, whereas all dislocation slip planes $\{011\}$ only have small angle orientation differences between adjacent martensitic laths. Thus, the effective grain size for cleavage fracture has been refined to an individual lath width, which is around a few hundreds of nanometers, whereas the effective grain size for plastic deformation is still at the micron level. As a result, we can improve toughness by grain size refinement without impacting strength.

This observation can be further applied to diffusional phase transformation in nanograined materials. In traditional bulk materials, most of the diffusional phase transformation caused by thermal and thermomechanical processes will be incoherent. However, for nanograined materials, coherent transformation, which will lead to a crystallographic relationship between adjacent domains, might be highly possible in order to minimize the interface energy. This is especially true in the case of the thermomechanical process when external elastic strain energy is involved. A similar phenomenon has been observed by different groups [14, 15] in thermomechanically treated ferritic steels. This significant difference between bulk and nanostructure materials will lead to some unique mechanical properties for nanograined structural materials; for example, materials may not lose their ductility at all when grain size reaches the sub-100-nm range. This has both practical and scientific implications: Practically, it will help materials scientists to determine which process to choose for one specific type of transformation and the corresponding crystallographic relationships in order to achieve the desired mechanical properties; scientifically, it will tell us the theoretical limit of strength and toughness for certain materials.

10.2 Nanoindentation

The indentation or hardness test has been widely used to determine the mechanical behavior of materials for many centuries due to its simplicity. Unlike other

mechanical tests, such as a uniaxial tensile test, the indentation hardness test requires no sample mounting and preparation and yet can still quantitatively determine a material's intrinsic resistance to permanent deformation.

In addition, unlike bulk materials, it is very hard to conduct traditional materials for many fewer dimension materials, such as thin films (2D), nanotubes (1D), or nanoparticles (0D) due to the difficulties of sample preparation. On the other hand, for the indentation test, the impacted volumes from test samples are proportional to the indenter size, making it a perfect candidate for mechanical testing for nanomaterials.

Nanoindentation was discovered and has received much attention in recent years, especially the in situ nanoindentation test under TEM. In this section, we first discuss the operating principles of nanoindentation and then introduce some examples of nanoindentation research.

10.2.1 Principles of Nanoindentation

In a traditional indentation hardness test, a tip with a hard material (usually diamond) is pressed down on a sample and held under a constant load P . After the load is removed, the indentation size A will be measured under a microscope and the hardness calculated through $H = P/A$. Clearly harder materials will have a smaller indentation size than softer materials under the same load using the same tip. Readers can find more details in any college introductory-level textbook on the mechanical behavior of materials, such as [16].

As mentioned earlier, the sample volume impacted by the indentation test is small and proportional to the indenter size. With today's advanced manufacturing technologies, we are able to make very small and precise indenter tips, down to the micron or even nanometer size, resulting in a very useful tool to characterize the mechanical behavior of thin films and nanomaterials.

When the indenter tip and sample size is small (down to nanometer size), very often it is difficult to measure the indentation size, even with an optical microscope. Instead, people have to record a load versus displacement (depth) curve in order to determine the indentation size (if the tip has a known shape with high precision) and other mechanical properties of the material being tested. A typical load P versus depth or displacement h curve is shown in Figure 10.3 including both loading and unloading. Many methods have been proposed to extract or evaluate a material's intrinsic mechanical properties from this curve, such as elastic modulus, stress/strain behavior, yield strength, and work hardening rate. Readers who are interested in this topic can find more details in [17].

10.2.2 In Situ Nanoindentation

It has been mentioned that nanoindentation is an excellent tool for characterizing the mechanical behavior of thin films and nanomaterials. However, it can only

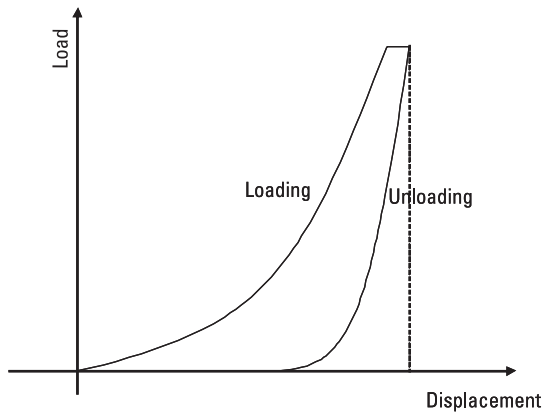


Figure 10.3 Typical load versus displacement curve for nanoindentation.

extract the mechanical properties of nanomaterials; it tells us very little about deformation mechanisms.

Recently, in situ nanoindentation experimental techniques have been introduced to conduct conventional nanoindentation tests with a TEM, allowing direct observation of nanoindentation-induced deformation behaviors. The experimental setup is shown in Figure 10.4 [18]. A special TEM sample holder was designed with a very sharp diamond tip [Figure 10.4(a)] that is connected with piezoceramics to control the load and record the displacement. The specimen is also prepared with a special geometry as a thin film deposited in a sharp silicon wedge produced by focused ion beam (FIB) etching so that the film is thin enough that electrons can transmit through it, allowing for direct observa-

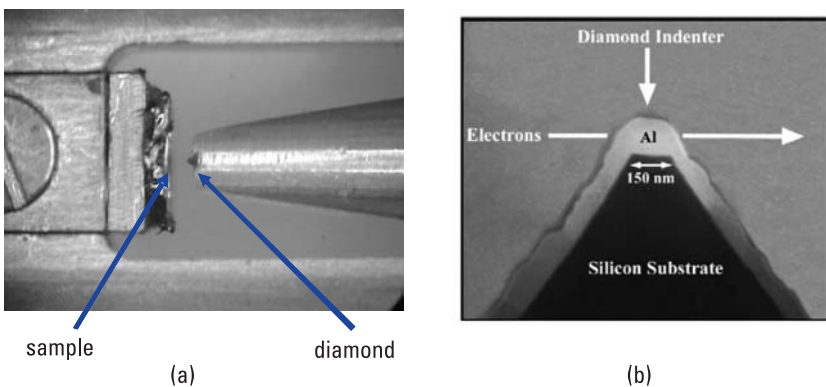


Figure 10.4 (a, b) Sample geometry for nanoindentation. (Reprinted with permission from [18]. Copyright 2004 Elsevier.)

tion. Using this technique, we can not only directly observe and record deformation behaviors during nanoindentation tests, but can also conduct diffraction pattern analysis in real time to characterize a material's crystal structures.

Dr. Andy Minor and his group at University of California–Berkeley and Lawrence Berkeley National Laboratory have conducted much research on nanostructural materials, such as nanograin-sized Al, nanoparticles such as CdS, and multilayer thin films. Here we merely introduce several examples [17–20].

10.2.2.1 Nanograin-Sized Al Films on Si Substrate

An in situ nanoindentation test is shown in Figure 10.4(b). In this experiment [18], a nanograin-sized pure Al film was deposited on a Si substrate with a wedge shape through an interrupt growth method. The film thickness is about 200 nm, and the grain size ranges from 50 to 200 nm so they are almost equiaxial nanograins.

Due to the small grain size, the grain boundary energy is high. Therefore, during the in situ nanoindentation test, stress-induced grain growth (or Oswald ripening) was observed. Figure 10.5 shows rapid grain ripening through grain boundary movement from a series of four images extracted from video clip recorded during an in situ nanoindentation test. The grain in the center of the image is about 150 nm and is located in the cusp of the grain boundary between three large grains. When the indenter starts to press the film, as shown in the sequence of images in Figures 10.5(a) through 10.5(d), the small grain dramatically shrinks or ripens by the neighboring larger grains, indicating that stress-induced Oswald ripening has occurred in order to minimize the high grain boundary energy of nanograin-sized Al film.

This is the first direct observation of the deformation mechanism of nanograin-sized structural materials.

10.2.2.2 CdS Nanoparticles

According to the Hall-Petch equation introduced at the beginning of the chapter, nanoparticles and nanograin-sized materials theoretically should offer us very high strength but typically very limited toughness or ductility. However, mechanical testing of a nanoparticle or nanotube can be nontrivial. A typical mechanical tensile test usually requires a large volume of the specimen, which is practically impossible for nanomaterials. Nanoindentation is, therefore, an ideal approach for characterization of the mechanical behavior of nanomaterials.

In one case [19], hollow CdS nanoparticles were tested using an in situ nanoindentation/compressive test under the observation of TEM. Figure 10.6 shows results from the direct observation of one CdS particle with a radius of 210 nm (shell thickness is about 30 nm) under in situ TEM [19]. Figure 10.6(a) is the image before any loading. In Figure 10.6 (b), the load has reached its maximum, yet the CdS is still in elastic deformation. In fact, the calculation

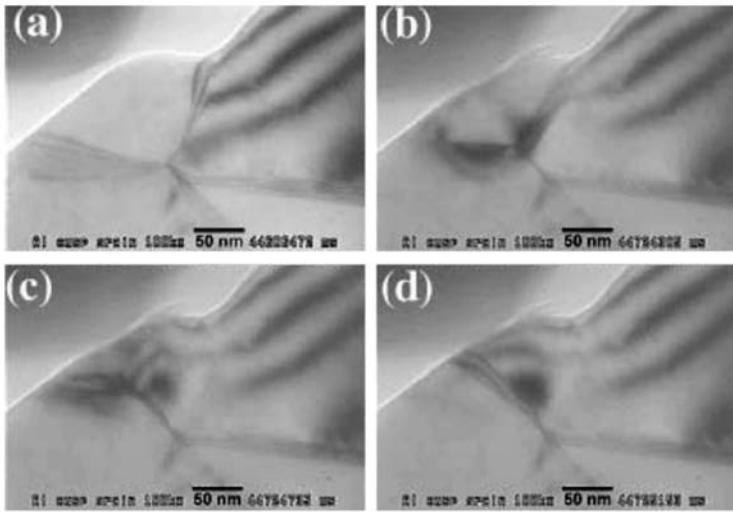


Figure 10.5 A series of images extracted from a videotape record showing stress-induced grain growth during an in situ nanoindentation test on nanograin-sized aluminum: (a) Before indentation the indenter approaches from the upper-left corner and (b) the indenter makes contact with a small grain in the cusp of three larger grains. (c) The small grain in the cusp starts to shrink and is left as a small film (d) between the neighboring grains. [Reprinted with permission from [18]. Copyright 2004 Elsevier.]

showed that at this point, the contact pressure was about 1 GPa and the shell was still capable of withstanding this extreme stress without any failure. This is approaching the ideal strength of CdS. In addition, this extremely high strength also enables the nanoparticles to exhibit considerably good deformation as shown in Figure 10.6(b). The elastic deformation can reach up to 20% before the nanoparticle breaks, as shown in Figure 10.6(c). Figures 10.6(d) and 10.6(e) show dark-field images of nanoparticles before and after the test, and Figure 10.6(f) shows the load versus displacement curve for this in situ mechanical test.

This test demonstrates that it is possible to achieve high strength and high ductility in the single nanoparticle during deformation. So it is a very valuable direct observation for characterization of mechanical properties of nanoparticles and nanograin-sized materials.

10.2.2.3 TiN/MgO Thin Films

Thin-film mechanics is another area that has attracted much attentions in recent years in an effort to understand the mechanical instability of thin films and nanostructures as part of a nanoscale electronic mechanical system (NEMS).

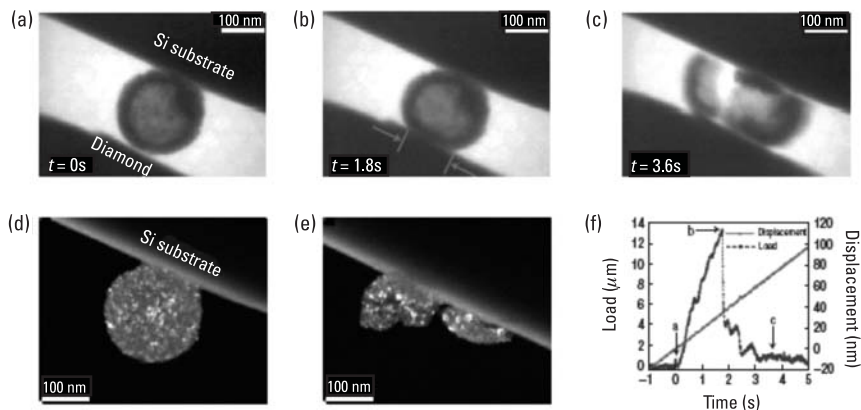


Figure 10.6 Example of a typical in situ compression test of an individual nanocrystalline hollow CdS sphere. (a–c) Extracted video frames of the in situ compression test, corresponding to (a) $t = 0\text{ s}$, (b) $t = 1.8\text{ s}$, and (c) $t = 3.6\text{ s}$. The estimated contact diameter is shown in (b). (d, e) Dark-field TEM images of the hollow nanocrystalline CdS ball resting on the Si substrate (d) before and (e) after the compression test. (f) Load and displacement data from the loading portion of the in situ test versus time. (Reprinted with permission from [19]. Copyright 2008 Nature Publishing Group.)

However, it may not be trivial to test the mechanical behavior of thin films, especially when film thickness goes down to the nanometer region. Nanoindentation is one of the very effective methods.

In [20], researchers from UC Berkeley and the National Center for Electron Microscopy (NCEM) at the Lawrence Berkeley National Lab used in situ nanoindentation to make direct observations on an epitaxy TiN film grown on a MgO (001) substrate. The sample had a geometry similar to that shown in Figure 10.4(b) through FIB milling techniques. Because TiN is a transition metal nitride with very high melting temperature, the nominal hardness is around 20 GPa, much higher than the hardness of the MgO substrate ($\sim 9\text{ GPa}$). So in the case of a hard/brittle thin film on a soft substrate, it is generally agreed that the soft substrate will experience substantial deformation during the indentation. As shown in Figure 10.7 [20], after indentation, a large amount of dislocation was generated and propagated through the MgO substrate, causing extensive plastic deformation. On the other hand, although limited, the TiN film also showed a small amount of dislocation, mainly around the plastic zone near the point of contact of the indentation. This is the first direct evidence proving that TiN films can exhibit some amount of plastic deformation even at room temperature. In addition, the film also buckled, causing a small angle of rotation on the axis of indentation near the contact region. Many

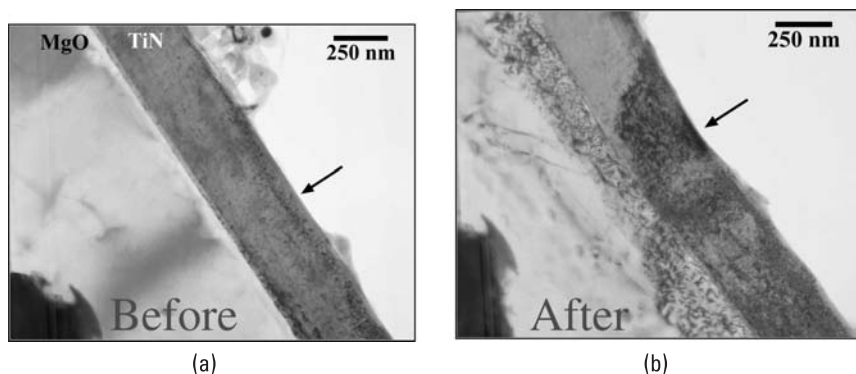


Figure 10.7 (a, b) TEM images from an in situ nanoindentation test on a TiN-MgO thin-film structure showing that the TiN buckled after indentation, indicating that a mechanical instability occurred on a hard film (TiN) on a soft substrate. (Reprinted with permission from Minor et al. (2003). Copyright 2003 Springer.)

dislocations formed in the plastic zone are also geometrically necessary dislocation for this small angle rotation. This buckling is evidence of mechanical instability, which we discuss extensively in the next section.

10.3 Mechanical Instability of Nanostructures

Polymeric nanostructures and thin films composed of elastic materials are critical for optical gratings, sensor arrays, actuators, biological fluid channels, and embossing/imprinting masks due to their ease of processing, exceptional chemical/optical properties, and large mechanical flexibility. However, when the dimensions of these structures are reduced to the nanometer scale or structures become tall and narrow, structural instability or buckling becomes a serious issue that could significantly affect the operation or reliability of devices, sensor elements, or nanomanufacturing processes. In this section, we introduce a few buckling-related phenomena and present some mechanics models corresponding to them.

10.3.1 Wrinkling of Thin Films

Buckling or wrinkling occurs in numerous forms across wide length scales; their common forms include the wrinkling of human skin, the surface of many dried fruits, and even the formation of mountain ridges. This wrinkling phenomenon has often been seen at a reduced scale by material scientists in the field of nanofabrication. In one particular example, plasma treatment of a soft polysiloxane surface generated a rigid add-layer, leading to the formation of wrinkles on

the top surface after the bottom portion was compressed. In another scenario of buckling, a thin metal film was deposited atop an elastic substrate; subsequent cooling introduced mismatch of shrinking between the two materials. Accordingly, biaxial compression was applied to the elastomer, which gave rise to the formation of a herringbone type of wrinkling. Some of these wrinkling and buckling phenomena are shown in Figure 10.8 [21].

The wrinkling patterns shown in Figure 10.8 were induced by thermal stress. Capillary wrinkling of floating, thin polymer films also has been reported [22]. A freely floating polymer film, tens of nanometers in thickness, wrinkles under the capillary force exerted by a drop of water placed on its surface (see Figure 10.9). In that report, films of polystyrene were spin-coated onto glass substrates. The film thickness h was varied from 31 to 233 nm. A circle of diameter of $D = 22.8$ mm was scribed onto the film with a sharp edge. When

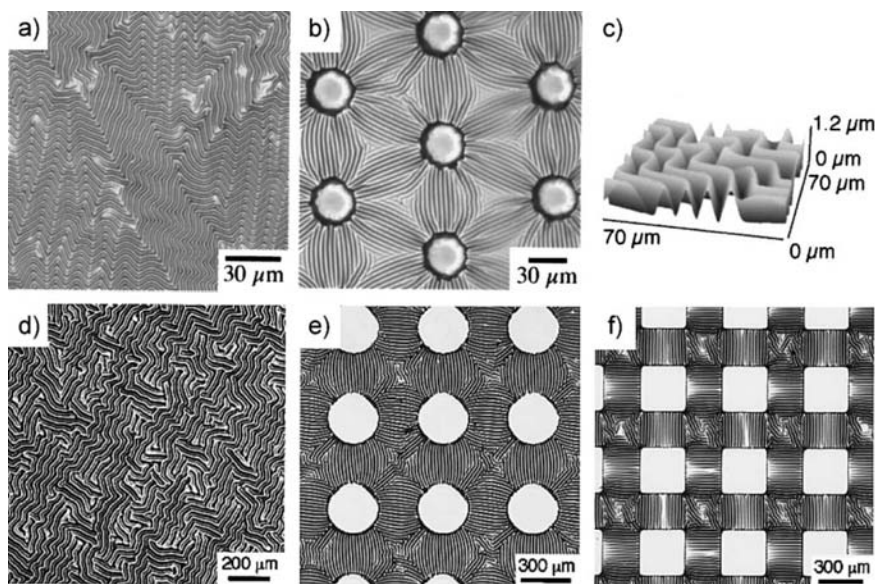


Figure 10.8 Optical micrographs of buckled surfaces prepared by plasma oxidation of a heated poly(dimethylsiloxane) (PDMS) sheet comprising (a) a homogeneous PDMS layer and (b) a PDMS substrate with posts (height: $5\text{ }\mu\text{m}$; diameter: $30\text{ }\mu\text{m}$) separated by $70\text{ }\mu\text{m}$. The buckles were formed upon cooling the sample to room temperature. (c) Scanning force microscopy image of disordered buckling waves. Optical micrographs of patterns formed when a thin layer of gold was deposited onto warm PDMS and the sample was cooled to room temperature. (d) Disordered patterns, (e) circles (radius: $150\text{ }\mu\text{m}$), and (f) flat squares (sides: $300\text{ }\mu\text{m}$) elevated by 10 to $20\text{ }\mu\text{m}$ relative to the surface showed ordered patterns of waves on the recessed regions and buckling on the plateau. (Reprinted with permission from [21]. Copyright 2006 Royal Society of Chemistry.)

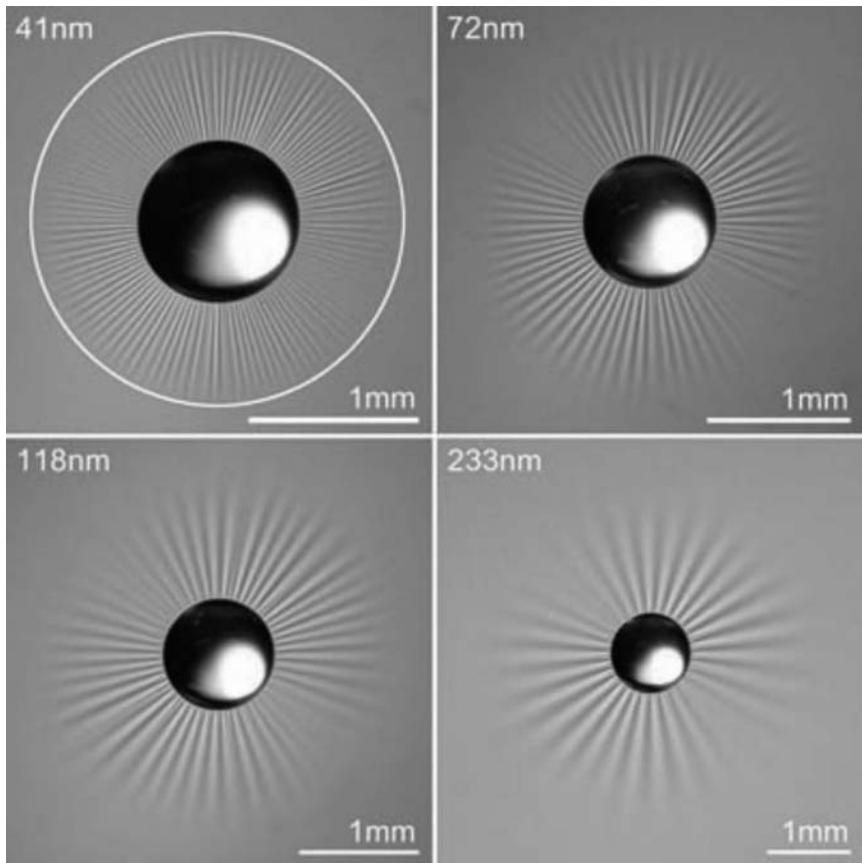


Figure 10.9 Four PS films of diameter $D = 22.8$ mm and of varying thicknesses floating on the surface of water, each wrinkled by water drops of radius $a \approx 0.5$ mm and mass $m \approx 0.2$ mg. As the film is made thicker, the number of wrinkles N decreases (there are 111, 68, 49, and 31 wrinkles in these images), and the length L of wrinkles increases (L is defined as shown at top left, measured from the edge of the water droplet to the white circle). The scale varies between images, whereas the water droplets are approximately the same size. (Reprinted with permission from [22]. Copyright 2007 American Association for the Advancement of Science.)

the substrate was dipped into a petri dish of distilled, deionized water, a circular piece of the PS film detached from the substrate. Because PS is hydrophobic, the film floated to the surface of the water where it was stretched flat by the surface tension γ of the air–water interface at its perimeter. Wrinkles were induced in the stretched, floating film by placing a drop of water in the center of the film (Figure 10.9), by placing a solid disk in the center of the film, or by poking the film with a sharp point to produce a fixed out-of-plane

displacement. All of these methods of loading led to qualitatively similar wrinkling patterns, radiating from the center of the load.

The wrinkling induced in Figure 10.9 is primarily due not to the weight of the drop, but to the capillary force exerted on the film by the surface tension at the air–water–PS contact line. The radial stress σ induced at the edge of the drop is dominated by the surface tension, which for the conditions of Figure 10.9 is about 100 times as great as the radial stress developed due to the weight of the drop ($mg/2\pi a$), where m is the mass of the drop and a its radius. Indeed, a solid object of weight and contact area comparable to those of the drops shown in Figure 10.9 would produce no discernible wrinkling.

To fully appreciate the wrinkling phenomenon in thin films, both a 1D mechanical model and a 2D model have been used. In the 1D model, the structure is simply considered to be an infinitely long film perfectly attached to the surface of an elastic semi-infinite medium (see Figure 10.10). The film and the medium are of width b in the y direction, where b is small; therefore, the system is in a state of plane stress ($\sigma_y = 0$). The shear stress between the upper layer and lower substrate is neglected. The cross section of the film is supposed to be a rectangular section of thickness t [23]. When the structure buckles, we can assume the surface resembles a sinusoidal wave with a half-wavelength l . Then with Timoshenko's theory of elastic stability [24], the following relationship can be obtained:

$$\frac{l}{t} = \pi \left[\frac{(3 - \nu_c)(1 + \nu_c)}{12} \right]^{1/3} \left(\frac{E}{E_c} \right)^{1/3} \quad (10.4)$$

where E_c and ν_c are the modulus of elasticity and Poisson's ratio for the substrate, respectively, and E is the modulus of elasticity of the film. Equation (10.4) states that the wavelength of the wrinkles depends only on the material properties of the sheet and the thickness of the thin film, and is independent of the applied stress and strain. The 1D model has been widely used to solve wrinkling problems

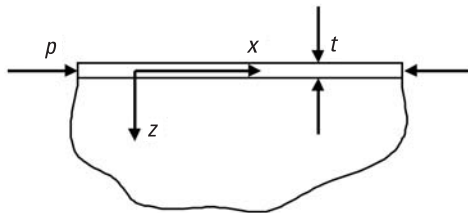


Figure 10.10 An infinitely long film attached to the face of an elastic semiinfinite medium.

[25–28]. Although it is simple, it predicts the direct proportionality between the wavelength and the thickness of the coating, where an increase in wavelength will appear when substrate rigidity decreases.

A 2D model has also been used to solve the wrinkling problem (see Figure 10.11). If we assume one of the wavelengths is infinite, the 2D model can be used to reduce the result for the 1D plane-strain model, which has been widely used to solve wrinkling problems [29, 30]. Associated with different critical buckling stresses, the covering film has different modes. Specifically, when the wavelengths in two directions are the same, that is, when $\lambda_1 = \lambda_2$, the wrinkles follow a square checkerboard mode, which is investigated in [31] by using the finite element method. Based on this 2D model, Song et al. [32] have established an analytic approach to the study of wrinkles having 1D, checkerboard, and ordered herringbone buckling patterns.

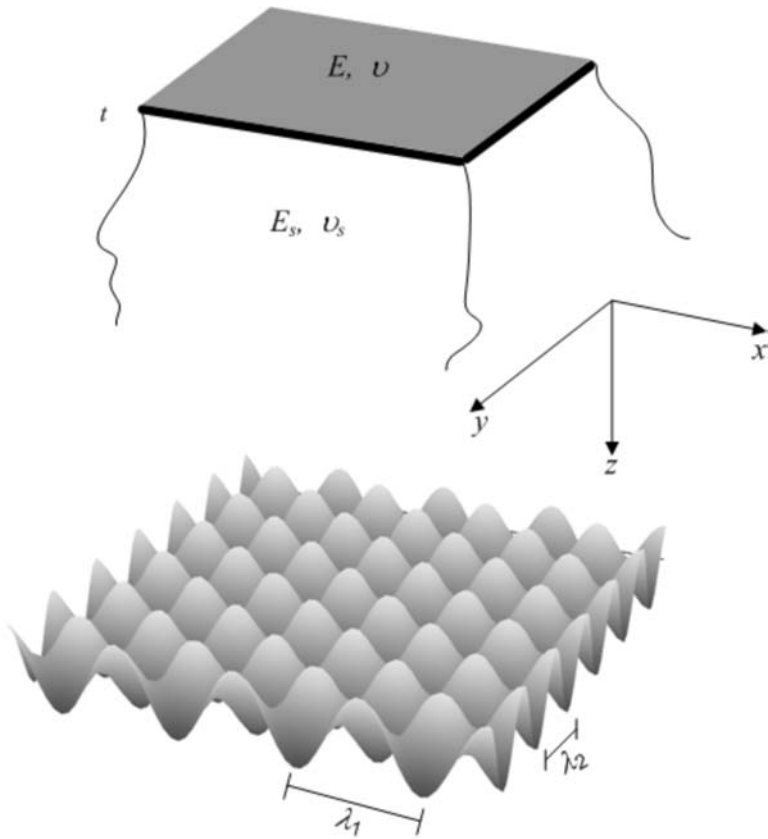


Figure 10.11 Schematic illustration of the geometry of wrinkles in a thin film residing on top of a thick elastic foundation.

10.3.2 Buckling of Spheroidal Core/Shell Structures

Spontaneous buckling of thin films on compliant substrates can achieve numerous highly ordered patterns due to mismatched deformation, which can be manipulated in different ways. However, most previous studies focused on the buckling arising in the planar geometry, and the effect on the curvature of the substrate was untouched. It is well known that the buckling of thin shells is highly sensitive to the shell's curvature. In addition, the buckling behavior on a closed surface will differ from that on a surface with free boundaries, since they are topologically distinct geometrical objects: The closed 2D surface such as a sphere has genus 0, whereas for a surface of free boundaries, planar or curved, it has genus 1. Therefore, the buckling patterns on a sphere will manifest some features primarily arising from the topological constraint. Li et al. [33] reported that the triangular patterns can self-assemble on the surface of a SiO_2 shell on a spherical Ag core structure by cooling. Various global pattern features can be reproduced by anisotropic stress-driven buckles on spheroidal core/shell systems. Three dimensionless parameters—the ratio of effective size/thickness, the ratio of equatorial/polar radii, and the ratio of core/shell moduli—primarily govern the initiation and formation of the patterns.

The numerical and experimental study of a spherical Ag core/ SiO_2 shell system has been studied (Figure 10.12) [34]. The transition of buckling patterns is observed when either the curvature of the system or the stress developed by

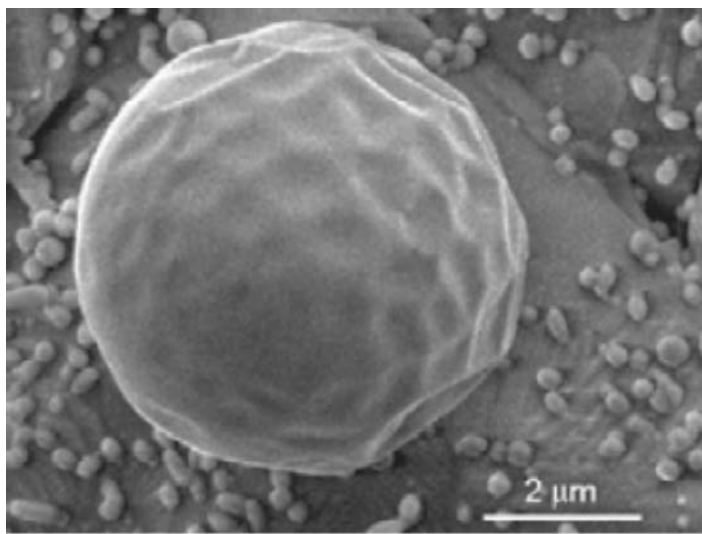


Figure 10.12 SEM micrographs on an Ag core/ SiO_2 shell system. (Reprinted with permission from [34]. Copyright 2008 American Physical Society.)

cooling is changed. The triangular dent-like buckles are preferred in smaller particles (due to the strong effect of curvature), whereas the labyrinth-like pattern dominates the larger particles. When the substrate radius is less than about 50 times the film thickness, the substrate curvature strongly influences the nominal film stress as well as the critical buckling wavelength. The critical buckling stress is very sensitive to the underlying substrate curvature and can be significantly reduced by enlarging the curvature of the core/shell structure. When the core size gets very big, all numerical solutions converge to that for a planar film bound on a semiinfinite substrate. The parallel experiments have verified the dominance of the triangularly distributed dent-like patterns in smaller particles and the formation of the labyrinth-like patterns in larger particles. Their results serve as a first attempt at controlling thin-film buckling patterns via the variation of substrate curvature, which may promise potential applications in a variety of areas.

10.3.3 Buckling of Nanobeams

Lateral collapsing of structures is another important mode of instability, which has been frequently seen among periodic arrays of nanostructures. To give readers a visual picture of the buckling in soft nanostructures, we used the easy-to-perform process of embossing/imprinting lithography and created periodic nanostructures atop an elastic material, PDMS. The rigid imprint stamp has 1D grating features with a height of 150 nm and a pitch of 200 nm [see Figure 10.13(a)]. The resulting PDMS copies show extensive surface buckling features as indicated by Figures 10.13(b, c). Although the results from the buckling analysis in structural engineering are not directly applicable to this case for the most part, there are results that are relevant and more or less shed some light from different angles; for example, the contact of two buckled columns, the buckling of a beam-like structure on an elastic foundation, and the lateral torsional buckling of a high beam with $h \gg 2a$. Some relevant previous knowledge on buckling of beam-like structures is shown in Figure 10.14.

Eulerian theory was one of the earliest theories for determining the critical load of buckling in structures other than films. It was derived by mathematician Leonhard Euler in 1757. It is obtained by considering the behavior of an ideal column. The model gives the maximum axial load that a long, slender, ideal column can carry without buckling. The ideal column is one that is perfectly straight, homogeneous, and free from initial stress. Also, it is based on the assumption that the cross sections, after deformation, remain planar and normal to the bent axis of the beam and also postulates a linear strain distribution across the cross section and ignores the influence of transverse shear deformations. Here, the boundary condition that is free at the upper end is taken as the derivative example.

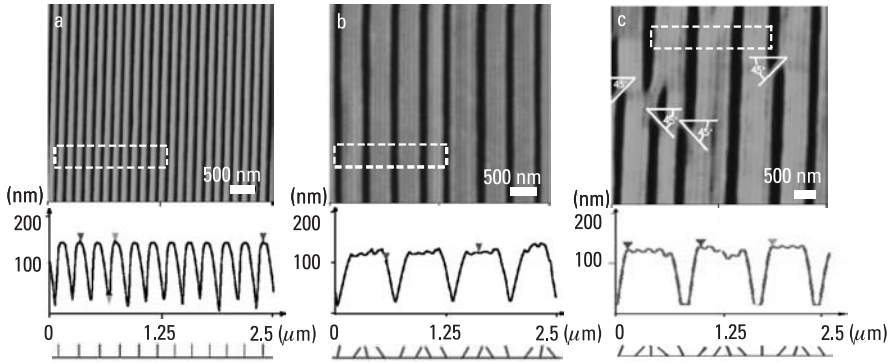


Figure 10.13 AFM topographical images and cursor plots of a rigid mold and fabricated PDMS structures from embossing/imprinting: (a) original mold with a pitch of 200 nm and a depth of 150 nm; (b) periodic buckling appears as multibeam pairing; and (c) both multibeam and multimode buckling features. Rotations of individual beams at selected areas are also sketched underneath each cursor plot: buckling following #10 mode in (b) and #12 mode in (c) as predicted by Figure 10.12. A branching angle of 45° is evident in the boundaries between the two different buckled patches. Note the existence of a slightly larger branching angle for pairing beams in the boundary.

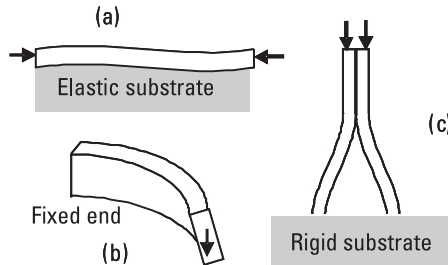


Figure 10.14 Relevant previous knowledge on buckling of beam-like structures: (a) a thin beam on an elastic foundation; (b) lateral torsional buckling of a high beam; and (c) buckling of two beams in contact.

When the column is assumed to slightly deflect under the vertical load P , the bending moment at any cross section (see Figure 10.15) is

$$M = -P(\delta - y) \quad (10.5)$$

Substitute (10.5) into the differential equation of Beam-columns

$$M = -EI \frac{d^2 y}{dx^2}, \text{ we obtain}$$

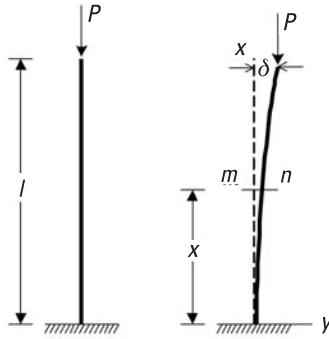


Figure 10.15 Buckling of a beam with a concentrated axial load.

$$EI \frac{d^2 y}{dx^2} = P(\delta - y) \quad (10.6)$$

With the boundary condition

$$\frac{dy}{dx} = 0, y = 0 \text{ at } x = 0 \quad (10.7)$$

$$y = \delta \text{ at } x = l \quad (10.8)$$

we get the critical load

$$P_{cr} = \frac{\pi^2 EI}{(kl)^2} \quad (10.9)$$

where E is the modulus of elasticity, I is the area moment of inertia, l is the unsupported length of the column, and k is the column effective length factor, whose value depends on the conditions of end support.

Euler's column theory takes various boundary conditions into consideration and also describes different buckled states under corresponding critical loads. His theory provides great help in solving engineering problems. However, in practice, the situations are more complicated. For example, the buckled problem cannot accurately be described as a small deformation, the buckled process is not elastic, and not all the forces applied on the buckled beams are vertical axial forces. All of these issues can be factors that influence the application of Euler's theory. To solve more complex problems in reality, Timoshenko [24] uses the energy method to describe problems. He also developed a theory of

buckled beams by considering more complicated situations, such as studying the buckling of a bar with intermediate compressive forces and the buckling of a bar under distributed axial loads.

In this case, the beam is buckled under longitudinally distributed loads (see Figure 10.16), which means the differential equation of the deflection curve of the buckled beam is no longer an equation with constant coefficients. Here, Timoshenko applied the infinite series to solve the problem. Similar to (10.6), the differential equation of the deflection curve is

$$EI \frac{d^2 y}{d^2 x} = \int_x^l q(\eta - y) d\xi \quad (10.10)$$

where q is the uniformly distributed load of intensity. Differentiating (10.10), we get

$$EI \frac{d^3 y}{dx^3} = -q(\eta - x) \frac{dy}{dx} \quad (10.11)$$

With the boundary conditions, we obtain

$$u = C_1 z^{-1/3} \left(1 - \frac{3}{8} z^2 + \frac{9}{320} z^4 - \dots \right), \quad z = \frac{2}{3} \sqrt{\frac{q}{EI}} (l - x)^3 \quad (10.12)$$

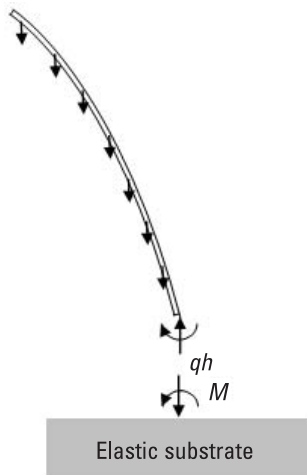


Figure 10.16 Buckling of a beam on an elastic substrate.

The critical value of the uniform load for the beam is then

$$(ql)_{cr} = \frac{7.837EI}{l^2} \quad (10.13)$$

The Eulerian beam model is usually employed to analyze the buckling behavior of beam-like nanostructures [35, 36]. In the experiments in [37], the mechanical properties of nanotubes were investigated based on the classical continuum model. Also, according to Euler's buckling model described by the continuum model, Young's moduli of the nanotube can be evaluated [38]. Based on nonlocal Timoshenko beam theory, buckling analysis of micro- and nanorods and tubes is investigated in [39].

10.3.4 Collective Buckling Model for Periodic Array of Nanostructures

Mechanical instability in a periodic array of a nanostructure presents a unique type of buckling problem associated with polymers during nanomanufacturing processes such as nanoimprint lithography or soft lithography. The materials of both the structures and the substrate are compliant and flexible (nanoimprint lithography uses rigid Si as a substrate, which may undergo substantial deformation under high loads) and are prone to buckling. In particular, the substrate could deform easily and, therefore, the thin film or neighboring structures could interact through the deformation of the substrate. Neighboring structures may also buckle into contact with each other. These unique features make the problem rather different from the well-studied buckling of elastic structures in civil, mechanical, and aerospace engineering.

To focus on the interaction, as an approximation we neglect the bending of the beams; buckling of a beam is qualitatively represented by the rotation of the beam. Our theoretical model consists of many interacting rigid beams on an elastic substrate [see Figure 10.17(a)]. The weight of a beam is represented by a vertical force P acting at the center of the beam. We consider rotations of the beams ϕ and their interactions with the substrate through couples or moments denoted by M . Figure 10.17(b) shows the notation for rotations of the substrate normals and surface couples.

Because the major deformation is rotation, as approximations we have neglected the horizontal and vertical displacements of the base of the beams. Within the linear theory of elasticity, by superposition, we can write

$$\phi_i = \sum_{j=1}^{\infty} \alpha_{ij} M_j \equiv \alpha_{i(i-1)} M_{i-1} + \alpha_{ii} M_i + \alpha_{i(i+1)} M_{i+1} \quad (10.14)$$

where the coefficients α_{ij} are called influence coefficients, physically representing the rotation of the normal at the i th location due to a unit couple at the j th

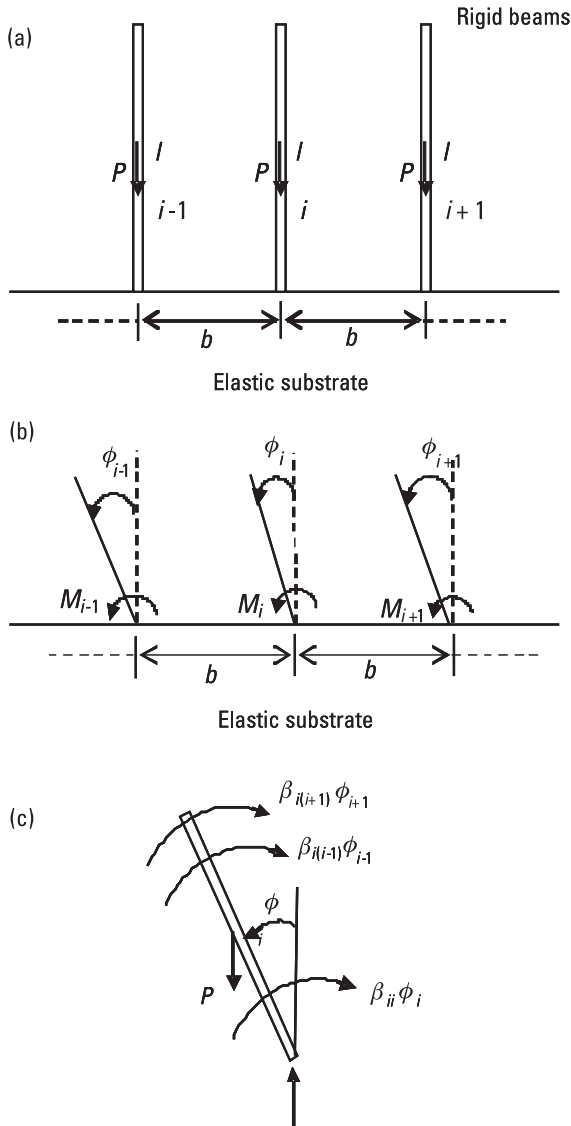


Figure 10.17 (a) A system of interacting rigid beams on an elastic substrate; (b) an elastic substrate with surface couples and rotations at the surface; and (c) a free body diagram of a buckled beam.

location only, with all other couples being zero. Equation (10.14) can be interpreted as describing rigid beams pinned at their base to a substrate with rotational springs for reactions from the substrate and interactions among beams. The α_{ij} terms then represent effective spring constants and can be calculated

from the theory of elasticity. In (10.14) we have made another approximation by considering interactions only between a beam and its two immediate neighbors. Long-distance interactions among nonneighboring beams are neglected. To calculate $\alpha_{i(i-1)}$, α_{ii} , $\alpha_{i(i+1)}$, we apply $M_i = M$ only (see Figure 10.18).

The solution to the problem in Figure 10.18 was given in [40]. Assuming the half-space is fixed at $r = \infty$, we obtain the displacement field as

$$\begin{aligned} u &= -\frac{2M \sin 2\theta}{E\pi r} \\ v &= -\frac{M}{E\pi r} (1 - \nu) \cos 2\theta + \frac{M}{E\pi r} (1 + \nu) \end{aligned} \quad (10.15)$$

The surface normal corresponds to $\theta = 0$. Because the bases of the beams have a finite dimension of $2a$, we define α_{ii} as the following average of the rotations of the fibers under a beam:

$$\alpha_{ii} = \left(\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{v(a, \theta)}{a} d\theta \right)_{M=1} = \frac{1 + \nu}{E\pi a^2} \quad (10.16)$$

For the interaction coefficients we calculate the surface slope at a neighboring beam:

$$\alpha_{i(i-1)} = \alpha_{i(i+1)} = \left(\frac{\partial v}{\partial r} \right)_{\substack{\theta=\pi/2 \\ r=b \\ M=1}} = \frac{-2}{E\pi b^2} \quad (10.17)$$

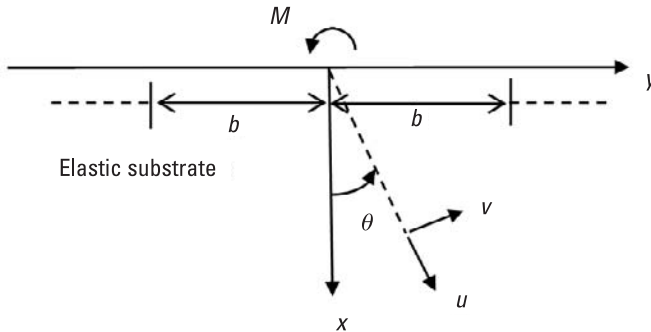


Figure 10.18 A half-space with one couple only.

For plane-strain problems, we need to make the following substitutions of the elastic constants in the above expressions:

$$E \rightarrow \frac{E}{1 - \nu^2}, \quad \nu \rightarrow \frac{\nu}{1 - \nu} \quad (10.18)$$

Equation (10.14) can then be inverted to yield

$$M_i = \sum_{j=1}^{\infty} \beta_{ij} \phi_j \cong \beta_{i(i-1)} \phi_{i-1} + \beta_{ii} \phi_i + \beta_{i(i+1)} \phi_{i+1} \quad (10.19)$$

where β_{ij} physically represents the moment at the i th location due to a unit rotation at the j th location only, with all other rotations being zero. The approximation in (10.19) is for nearest-neighbor interactions. In our numerical calculations, though, the full matrix of β_{ij} will be kept. Figure 10.17(c) shows the free body diagram of a buckled beam.

The equilibrium of the beam requires a zero moment about the base of the beam:

$$\beta_{i(i-1)} \phi_{i-1} + \left(\beta_{ii} - \frac{Ph}{2} \right) \phi_i + \beta_{i(i+1)} \phi_{i+1} = 0 \quad (10.20)$$

Equation (10.20) is a homogeneous system of linear equations. The trivial solution with all $\phi_i = 0$ is the unbuckled state. We are interested in nontrivial solutions representing buckled states; then (10.20) becomes an eigenvalue problem. We look for values of $Ph/2$ corresponding to which nontrivial solutions of ϕ_i exist.

Two types of energies are involved in the buckling process just discussed: strain energy and gravity potential. During the buckling process, the system gains strain energy but loses gravity potential energy. A buckled state represents a configuration of the system in which the strain energy gained is equal to the potential energy lost. In reality, a buckled state with less strain energy appears before a state with more strain energy. To order the buckled states, we only need to calculate the strain energy by integrating the strain energy density over the substrate or by using the effective spring constants. A more visual picture of all of the associated potential energies can be seen in Figure 10.19, indicating 20 modes of collective buckling in the Y - Z plane plus their eigenvectors together. Specifically, the first mode (#1) is an alternative picture of the eigenvector in Figure 10.20, and the rest of the modes show beam rotations due to their own specific eigenvectors. Because each individual mode of the collective buckling occurs at the expense of a certain amount of energy, two extreme cases among

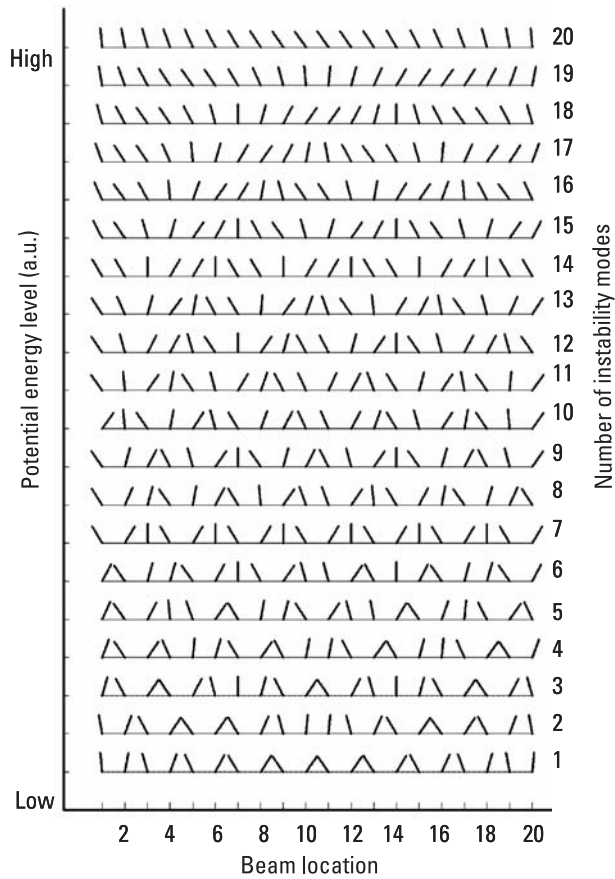


Figure 10.19 Multiple modes of collective buckling among 20 beams and their relationships with associated energy levels.

these 20 modes are worthy of emphasis: (1) When the system is affected with the lowest energy, periodic buckling between two neighboring beams will dominate the instability; and (2) all 20 beams rotated toward one direction will appear under the highest energy impact (#20). The first case of instability (1), that is, two-beam pairing, was discovered by many other reports [41–45] during microcontact printing and often used as a classic picture for nanostructure buckling. In those findings, a rigid substrate backing or a thick PDMS sample was often selected to limit large deformations to the stamp material [46]; therefore, a minimum energy impact was ensured for the occurrence of pairing. In contrast, when no special caution is given to the stamp material or a substantial deformation is involved in the process of nanostructuring, multibeam buckling modes at higher energy levels should be expected; fluctuation of those energy

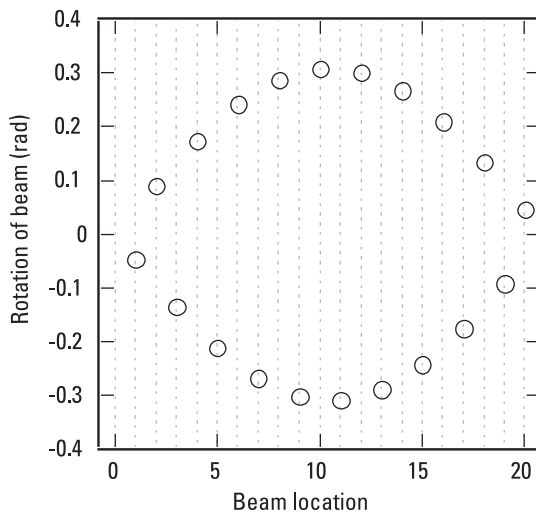


Figure 10.20 Eigenvector representing a collective buckled state of 20 beams with the lowest eigenvalue, where components of alternating signs represent neighboring beams rotating in opposite directions.

impacts could even generate multiple modes of buckling. A good match of the above predictions with experimental data is shown in Figures 10.13(b, c), where both multibeam and multimode buckling phenomena are evident. Particularly, selected areas in Figures 10.13(b, c) match with predicted #10 and #12 modes of buckling, respectively (sketched underneath each cursor plot in Figure 10.13). Furthermore, a branching angle of 45° between two neighboring buckling modes in Figure 10.13(c) suggests a shear or rotation nature for these impacts, validating the approximation of buckling in elastic beams with the rotation of rigid ones.

The preceding analysis also highlights an important question as to whether those buckling modes are selectable or can be controlled to some extent. Because buckling usually means failure in the field of nanolithography or device fabrication, existing strategies involve trying to secure the stability of nanostructures. Available means are mostly empirical and include replacing the nanostructure material with a tougher one [46–48] or a redesign of the geometrical shapes of the nanostructures [45]. Alternatively, the preceding analytical results suggest the likelihood of selectively promoting certain modes of nanostructure buckling by limiting associated energy levels. The benefit of doing so is that the number of buckling modes can be confined to a minimum level and the corresponding buckled features can be further utilized for nanolithography or other purposes. Because this strategy does not require the redesign of nanostructures with new geometry or materials, it could be applied as a universal means in broad applications of lithography, device fabrication, and design of sensors and actuators.

In the section, we reviewed the scientific aspects of wrinkling or buckling and the related phenomena. Until now, buckling-related research has focused on two issues: wrinkles brought about from the contraction of bilayer structures, and buckled beams resulting from lithography. We presented various models or theoretical foundations for the buckling.

References

- [1] Hall, E. O., *Proc. of the Physical Society B*, Vol. 64, 1951, p. 727.
- [2] Petch, N. J., *J. Iron Steel Inst.*, Vol. 174, 1953, p. 25.
- [3] Cottrell, A. H., *Trans. AIME*, Vol. 212, 1958, p. 192.
- [4] Petch, N. J., *Philosophical Magazine*, Vol. 3, 1958, p. 1089.
- [5] Griffith, A. A., *Proc. Int. Congr. Appl. Mech.*, Vol. 55, 1924.
- [6] Stollhoff, N. S., *Fracture*, Vol. 6, H. Liebowitz, (Ed.), New York: Academic Press, 1969, p. 7.
- [7] Kim, Y. H., Ph.D. Thesis, University of California, Berkeley, 1984.
- [8] Choo, W. Y., POSCO Research Laboratories, Pohang, Korea, Private Communication, 2001.
- [9] Nagai, K., National Research Institute for Metals, Tsukuba, Japan, Private Communication, 2000.
- [10] Morris, J. W., Jr., Guo, Z., Krenn, C., Kim, Y. H., *ISIJ International*, Vol. 41, No. 6, 2001.
- [11] Morris, J. W., Jr., J. I. Kim, and C. K. Syn, in *Advances in Metal Processing*, Burke, Mehrabian, and Weiss (Eds.), New York: Plenum Press, 1981, p. 73.
- [12] Morris, J. W., Jr., in *Mechanical Properties and Phase Transformations in Engineering Materials*, Antolovich, Ritchie, and Gerberich (Eds.), Warrendale, PA: TMS, 1986, p. 99.
- [13] Kim, H. J., Y. H. Kim, and J. W. Morris, Jr., *ISIJ International*, Vol. 38, 1998, p. 1277.
- [14] Guo, Z., C. S. Lee, and J. W. Morris, Jr., "On the Coherent Transformation of Steels," *Acta Materialia*, Vol. 52, No. 19, 2004, pp. 5511–5518.
- [15] Guo, Z., and J. W. Morris, Jr., "The Martensite Variant Generated by the Mechanical Transformation of Precipitated Interlath Austenite," *Scripta Materialia*, Vol. 53, No. 8, 2005, p. 933.
- [16] Courtney, T. H., *Mechanical Behavior of Materials*, Long Grove, IL: Waveland Press, 1996.
- [17] Gouldstone, A., et al., "Indentation Across Size Scales and Disciplines: Recent Developments in Experimentation and Modeling," *Acta Materialia*, Vol. 55, 2007, pp. 4015–4039.
- [18] Miao, J., et al., "Direct Observation of Deformation-Induced Grain Growth During Nanoindentation of Ultrafine-Grained Al at Room Temperature," *Acta Materialia*, Vol. 52, 2004, pp. 5381–5387.

- [19] Shan, Z. W., et al., "Ultrahigh Stress and Strain in Hierarchically Structured Hollow Nanoparticles," *Nature Materials*, Vol. 7, 2008, pp. 947–952.
- [20] Minor, A. M., et al., "In-Situ Nanoindentation of Epitaxial TiN/MgO (001) in a Transmission Electron Microscope," *J. of Electronic Materials*, Vol. 32, No. 10, 2003, pp. 1023–1027.
- [21] Genzer, J., and J. Groenewold, "Soft Matter with Hard Skin: From Skin Wrinkles to Templating and Material Characterization," *Soft Matter*, Vol. 2, 2006, pp. 310–323.
- [22] Huang, J. S., et al., "Capillary Wrinkling of Floating Thin Polymer Films," *Science*, Vol. 317, 2007, pp. 650–653.
- [23] Allen, H. G., *Analysis and Design of Structural Sandwich Panels*, New York: Pergamon, 1969.
- [24] Timoshenko, S. P., and J. M. Gere, *Theory of Elastic Stability*, New York: McGraw-Hill, 1961.
- [25] Bowden, N., et al., "Spontaneous Formation of Ordered Structures in Thin Films of Metals Supported on an Elastomeric Polymer," *Nature*, Vol. 393, No. 6681, 1998, pp. 146–149.
- [26] Volynskii, A. L., et al., "Mechanical Buckling Instability of Thin Coatings Deposited on Soft Polymer Substrates," *J. of Materials Science*, Vol. 35, No. 3, 2000, pp. 547–554.
- [27] Stafford, C. M., et al., "A Buckling-Based Metrology for Measuring the Elastic Moduli of Polymeric Thin Films," *Nature Materials*, Vol. 3, No. 8, 2004, pp. 545–550.
- [28] Efimenko, K., et al., "Nested Self-Similar Wrinkling Patterns in Skins," *Nature Materials*, Vol. 4, No. 4, 2005, pp. 293–297.
- [29] Kwon, S. J., "Wrenching of a Metal Thin Film on a Structured Polymer Layer," *J. of Physics: Condensed Matter*, Vol. 18, No. 41, 2006, pp. 9403–9413.
- [30] Hendricks, T. R., and I. Lee, "Wrinkle-Free Nanomechanical Film: Control and Prevention of Polymer Film Buckling," *Nano Letters*, Vol. 7, No. 2, 2007, pp. 372–379.
- [31] Chen, X., and J. W. Hutchinson, "Herringbone Buckling Patterns of Compressed Thin Films on Compliant Substrates," *J. of Applied Mechanics (Transactions of the ASME)*, Vol. 71, No. 5, 2004, pp. 597–603.
- [32] Song, J., et al., "An Analytical Study of Two-Dimensional Buckling of Thin Films on Compliant Substrates," *J. of Applied Physics*, Vol. 103, No. 1, 2008, 014303.
- [33] Li, C. R., X. N. Zhang, and Z. X. Cao, "Triangular and Fibonacci Number Patterns Driven by Stress on Core/Shell Microstructures," *Science*, Vol. 309, 2005, pp. 909–911.
- [34] Cao, G. X., et al., "Self-Assembled Triangular and Labyrinth Buckling Patterns of Thin Films on Spherical Substrates," *Physical Review Letters*, Vol. 100, 2008, 036102.
- [35] Harik, V. M., "Mechanics of Carbon Nanotubes: Applicability of the Continuum-Beam Models," *Computational Materials Science*, Vol. 24, No. 3, 2002, pp. 328–342.
- [36] Luo, C., A. Francis, and X. H. Lin, "Determination of Compressive Residual Stress in a Doubly Clamped Microbeam According to Its Buckled Shape," *Microelectronic Engineering*, Vol. 85, No. 2, 2008, pp. 339–347.

-
- [37] Wong, E. W., P. E. Sheehan and C. M. Lieber, "Nanobeam Mechanics: Elasticity, Strength, and Toughness of Nanorods and Nanotubes," *Science*, Vol. 277, No. 5334, 1997, pp. 1971–1975.
- [38] Mitsumasa, N., A. Seiji, and N. Yoshikazu, "Buckling Test Under Axial Compression for Multiwall Carbon Nanotubes," *Japanese J. of Applied Physics*, Vol. 44, No. 34, 2005, pp. L1097–L1099.
- [39] Wang, C. M., et al., "Buckling Analysis of Micro- and Nano-Rods/Tubes Based on Nonlocal Timoshenko Beam Theory," *J. of Physics D: Applied Physics*, Vol. 39, No. 17, 2006, pp. 3904–3909.
- [40] Timoshenko, S. P., and J. N. Goodier, *Theory of Elasticity*. New York: McGraw-Hill, 1970.
- [41] Delamarche, E., et al., "Stability of Molded Polydimethylsiloxane Microstructures," *Advanced Materials*, Vol. 9, No. 9, 1997, pp. 741–746.
- [42] Xia, Y. N., and G. M. Whitesides, "Soft Lithography," *Annual Review of Materials Science*, Vol. 28, 1998, pp. 153–184.
- [43] Hui, C. Y., et al., "Constraints on Microcontact Printing Imposed by Stamp Deformation," *Langmuir*, Vol. 18, No. 4, 2002, pp. 1394–1407.
- [44] Chuang, W. C., C. T. Ho, and W. C. Wang, "Fabrication of a High-Resolution Periodical Structure Using a Replication Process," *Optics Express*, Vol. 13, No. 18, 2005, pp. 6685–6692.
- [45] Lee, T. W., O. Mitrofanov, and J. W. R. Hsu, "Pattern-Transfer Fidelity in Soft Lithography: The Role of Pattern Density and Aspect Ratio," *Advanced Functional Materials*, Vol. 15(10) 2005. 1683–1688.
- [46] Schmid, H., and B. Michel, "Siloxane Polymers for High-Resolution, High-Accuracy Soft Lithography," *Macromolecules*, Vol. 33, No. 8, 2000, pp. 3042–3049.
- [47] Choi, K. M., and J. A. Rogers, "A Photocurable Poly(dimethylsiloxane) Chemistry Designed for Soft Lithographic Molding and Printing in the Nanometer Regime," *J. of the American Chemical Society*, Vol. 125, No. 14, 2003, pp. 4060–4061.
- [48] Rolland, J. P., et al., "High-Resolution Soft Lithography: Enabling Materials for Nanotechnologies," *Angewandte Chemie International Edition*, Vol. 43, No. 43, 2004, pp. 5796–5799.

About the Authors

Zhen Guo is a staff technologist at the Intel Corporation and a part time lecturer in the department of chemical engineering and materials science at San Jose State University. At Intel, he has been working on thin film deposition and process integration on 90 nm, 65 nm, 45 nm, and 32 nm flash and logic processes. He received his Ph.D. from the University of California, Berkeley in 2001; his M.S. from the University of Michigan, Ann Arbor in 1998; and his bachelor's degree in engineering from Tsinghua University in China; all in the field of materials science. His research covers a wide range of materials science, including ceramics, metallic alloys, Martensitic phase transformation, semiconductor materials, dielectric materials, and nanomaterials.

Li Tan is an assistant professor in the College of Engineering at the University of Nebraska–Lincoln (UNL). He is affiliated with the Engineering Mechanics Department's Materials Engineering Program and also holds appointments in the Department of Chemistry. His research interests are in the frontiers of nanofabrication and nanomaterials. His group builds everything from well-defined nanometers to molecular scale objects. Constructed, macroscopic structures not only inherit the personality of a nano-object, but also show unique electrical, thermal, or mechanical properties. His current projects include material design and surface engineering for organic field-effect transistors, material processing and manufacturing for organic photovoltaic devices, and mechanics of nanomaterials. His research has been supported by NSF, DoD, the state of Nebraska, and Agilent Technology, Inc. Before coming to UNL in 2005, Li worked as a research fellow at the Solid State Electronics Laboratory (U. Michigan) and as a visiting postdoctoral scientist at the Department of Chemistry (U. California). In 2002 he completed his doctoral studies at the Macromolecular

Science and Engineering Program of the University of Michigan-Ann Arbor. He also holds M.E. and B.E. degrees from Tsinghua University, China.

Index

- Adsorption, 67–68
 - chemical, 68, 69
 - defined, 67
 - illustrated, 67
 - physical, 68, 69
 - sites, 67–68
- AFM lithography, 114–19
 - advantages, 114
 - defined, 115
 - images, 116
 - nanografting, 117–19
 - nanindentation, 115–17
 - scratching, 115–17
 - thermochemical writing by, 117
- Air cushion press (ACP), 111–13
- Atomic bonding, 25–40
 - classic, 25–26
 - in crystalline solids, 35–40
 - four-atom molecule, 31–32
 - many-atom molecule, 34–35
 - in molecules, 26–35
 - six-atom molecule, 32–34
 - three-atom molecule, 31
 - two-atom molecule, 29–30
- Atomic force microscope (AFM), 87, 89
 - sharp probe tips with, 115
 - See also* AFM lithography
- Atomic layer deposition (ALD), 68, 104–5
- Auger electrons, 82–83
- Auger electron spectroscopy (AES), 90
- Backscattered electrons, 82
- Ballistic transport, 175
- Band structure
 - of CNT, 171–73
 - of CNT with armchair configuration, 172
 - 3D illustration, 171
- Band theory, 35–40
- Beams
 - with axial load, 227
 - bending of, 229
 - buckling of, 225–29
 - collective buckled state of, 234
 - on elastic substrate, 228
 - equilibrium, 232
 - multiple modes of collective buckling, 233
 - rigid, interacting, 230
 - uniform load, 229
- Benzene ring, 32–34
- Beryllium (Be) atoms, 38
- Biodiagnostics, 198–202
 - nanoparticle-based detection methods, 200–201
 - nanostructures in, 198–202
 - nanotube-based detection methods, 201–2
 - nanowire-based detection methods, 201–2

- Biomimetic nanotechnology, 188–98
 DNA, 188–91
 structural biomimicry, 191–98
- Blackbody radiation theory, 7
- Block copolymers, templated self-assembly (TSA) of, 122–26
- Body-centered cubic (BCC) structure, 50
 close-packed directions and planes, 51–52
 highest packing density, 52
 illustrated, 50
 packing density, 55
 surface structure, 56–57
- Bohr magnetron, 178
- Bonding
 atomic, 25–40
 length, 26
- Bone, 193, 194
- Bottom-up method
 for carbon-based nanocrystals, 44–47
 defined, 40
 See also Top-down method
- Bragg's law, 77–78
- Buckling
 beam-like structures, 226
 of beam on elastic substrate, 228
 of beam with axial load, 227
 of nanobeams, 225–29
 for periodic array of nanostructures, 229–35
 of spheroidal core/shell structures, 224–25
 thin films, 220
- C₆₀ fullerene ball, 46–47
- Carbon-based structures
 carbon nanotube, 45–47
 diamond, 44
 graphite, 45
- Carbon nanotubes (CNTs), 45–47
 band structure, 171–73
 C₆₀ fullerene ball, 46–47
 chiral angle, 84
 classifying, 45
 CVD methods, 173–75
 defined, 45
 fabrication of, 173–75
 as graphene, 69–70
 introduction to, 170–73
 multiwall (MWCNTs), 70, 173
 periodic boundary condition, 171
 PVD methods, 173
 schematic illustration, 46
 single wall (SWCNT), 70, 173
 zigzag, 45
 See also CNT-based nanoelectronic devices
- Cathode material
 current versus applied voltage, 8
 obtaining work function of, 9
- CdS nanoparticles, 216–17
- Cell study nanostructures, 202–6
 extracellular matrix (ECM), 202
 microarray platform, 202–6
- Chemical adsorption, 68, 69
- Chemical vapor deposition (CVD), 102–5, 168
 atomic layer deposition (ALD), 104–5
 CNT fabrication, 173–75
 deposition rate, 102
 normal reactant, 103–4
 plasma-enhanced (PECVD), 104, 168
 process illustration, 103
 summary, 105
 See also Thin-film deposition
- Chiral vectors, 45
- Classic atomic bonding, 25–26
- Close-packed structures, 54–55
 BCC structure versus, 55
 surface structure, 55–56
- CNT-based nanoelectronic devices, 170–77
 CNT interconnect, 175–76
 CNT transistor, 176–77
 See also Carbon nanotubes (CNTs)
- CNT interconnects, 175–76
- CNT-MOSFETs, 176–77
 defined, 176
 fabrication, 177
 top gate structure, 177
- CNT transistors, 176–77
- Compound eyes, 196–98
 design complexity, 196
 natural versus artificial comparison, 197
 in physiological optics, 196
- Compton's X-ray scattering experiment, 10
- Conduction band (CB), 37, 40
- Constructive diffraction, 76
- Coprecipitation, 130–32
- Coulomb blockade, 160
 conditions, 161

- excessive charging energy, 164, 165
 - region, 162
 - satisfying, 161
 - tunneling and, 163
- Coulomb force, 25
- Coulombic potential energy, 21
- Coulomb integrals, 27, 28
- Coulomb oscillation, 162, 165
- Coulomb staircase, 158
- Crystalline solids, atomic bonding in, 35–40
- Crystallographically preferred surface, 59
- Crystallography, 50–51, 55–57
- Crystal structures, 49–50
- Datta-Das SFET, 181
- Dc sputtering, 101–2
- De Broglie relationship, 16–17
- De Broglie's hypothesis, 11–12
- Destructive infringement, 76
- Dewetting, 73
- Diamond, 44
- Diamond structure, 39
- Diffraction, 75–80
 - constructive, 76
 - defined, 75–76
 - occurrence, 76
- Dip-pen nanolithography (DPN), 119–20
- Dispersion relationship
 - defined, 13
 - between energy and wave number, 17
- DNA
 - complementary probes, 202
 - complementary target, 201
 - fluorophore-labeled, 200
 - label-free detection of, 201
 - linear double helices, 188
 - mutant, 201
 - nanotechnology, 188–91
 - self-assembly features, 188
 - system control, 189
 - target, 201, 202
 - tile-mediated self-assembly, 189
 - tile system design, 190
- Drug delivery, 206
- Duality of electrons, 10–17
- Duality of light, 9–10
- Einstein, Albert, 4, 5, 7–9
- Electron energy loss spectrum (EELS), 86
- Electron microscopy, 80–89
 - fundamental principle, 81
 - interaction between electron beams and solids, 81–83
 - SEM, 86–87
 - SPM, 87–89
 - TEM, 83–86
- Electrons
 - Auger, 82–83
 - backscattered, 82
 - double-slit interference, 11
 - duality of, 10–17
 - free, 16–17
 - kinetic energy of, 8
 - in potential well, 17–21
 - secondary, 82
 - thin metal foil diffraction, 11
 - transmitted, 82
 - wave character, 12
 - as waves, 11–12
- Electrospinning, 206
- Electrospun nanofibers, 206
- Electrostatic assembly, 141–42
- Energy bands
 - conduction band (CB), 37, 40
 - continuous, 37
 - defined, 36
 - for insulators and semiconductors, 38–40
 - partially filled, for metals, 37–38
 - in solids, 36–37
 - valance band (VB), 36–37, 40
- Energy-dispersive X-ray spectroscopy (EDX), 87
- Eulerian beam model, 229
- Eulerian theory, 225, 227
- Excessive charging energy, 156
- Extracellular matrix (ECM), 202
- Face-centered cubic (FCC) structure, 50
 - atomic packing factor, 53
 - close-packed directions and planes, 53
 - illustrated, 50
 - packing density, 54
 - surface structure for, 55–56
 - See also* Close-packed structures
- Face-selective assembly, 143
- Field-effect transistors (FETs)
 - single-electron capacitors versus, 162
 - spin (SFET), 181–82
- Focused ion beam (FIB), 84, 215
- Four-atom molecule, 31–32
- Frank-van der Merwe model, 97

- Free electrons, 16–17
 kinetic energy, 16
 Schrödinger equation, 16
Full width at half-maximum (FWHM), 80
- Gecko foot, 191–93
 adhesive system, 192
 setae, 191
- Giant magnetoresistance (GMR)
 effect, 178–80
 storage device, 180–81
- Gibbs free energy, 71
- Grain refinement, 210–13
 approaches to, 210–13
 lath martensitic steel, 211
 reason for, 210
 thermal mechanical processing, 212–13
 thermal method, 213
- Graphene
 band structure of, 170–71
 CNTs as, 69–70
 as “gapless” semiconductor, 171
- Graphite, 45, 170–71
- Griffith theory, 210
- Hall-Petch equation, 210
- Hall-Petch slope, 210
- Helium atom, 22–23
 defined, 22
 energy dependence, 23
 potentials, 22–23
- Heterogeneous film growth mechanisms,
 97–99
 Frank-van der Merwe model, 97
 illustrated, 98
 Stranski-Krastanov model, 98–99
 Volmer-Weber model, 97–98
 See also Thin-film deposition
- Hexagonal close packed (HCP) structure, 50
 close-packed directions and planes, 54
 illustrated, 50
 surface structure for, 55–56
 volume packing density, 54
 See also Close-packed structures
- Highest occupied molecular orbitals
 (HOMOs), 32
- High-resolution TEM (HRTEM), 84, 134
 embedded nanoparticles, 169
 postanneal films, 169
- Homogeneous film growth mechanisms,
 95–97
 illustrated, 96
 island growth, 97
 step propagation, 96–97
 See also Thin-film deposition
- Hund’s rule, 24
- Hydrogen atom, 21–22
 Coulombic potential energy, 21
 quantum numbers, 22
- Hydrogen bonding-based assembly, 140–41
- Hydrophobic assembly, 144
- Hydrothermal/solvothermal methods,
 136–37
- Indentation hardness test, 213–14
- Influence coefficients, 229
- In situ nanoindentation, 214–19
 CdS nanoparticles, 216–17
 nanograin-sized Al films on Si substrate,
 216
 TiN/MgO thin films, 217–19
 See also Nanoindentation
- Insulators, energy band for, 38–40
- Island growth, 97
- Kinetic energy (KE), 7
 of emitted electrons, 8
 free electrons, 16
 measuring, 9
- Lateral surface relaxation, 63
- Lattices
 defined, 49
 groups of, 49
 structures, 50
- Laue method, 76–77
- Li atom, 37–38
- Light, 5–6
 duality of, 9–10
 particle nature of, 9
- Linear combination of atomic orbitals
 (LCAO) approximation
 basis, 26
 defined, 26
 four-atom molecule, 31–32
 many-atom molecule, 34–35
 six-atom molecule, 32–34
 theory, 26–35
 three-atom molecule, 31
- Lorentz force, 179
- Lowest unoccupied molecular orbitals
 (LUMOs), 32

- Magnetoresistive random access memory (MRAM), 182–84
advantages, 183
chips, 184
defined, 182
operating principle, 182
schematic illustration, 183
- Many-atom molecule, 34–35
- Maxwell's field equations, 5
- Mechanical instability, 219–35
buckling of nanobeams, 225–29
buckling of spheroidal core/shell structures, 224–25
for periodic array of nanostructures, 229–35
thin film wrinkling, 219–23
See also Nanostructures
- Metal-oxide semiconductor field-effect transistors (MOSFETs), 154, 176–77
- Metals, partially filled energy band for, 37–38
- Microarray platform
cell culture, 203
as research tool, 202–6
via sequential illumination, 203
- Microcontact printing, 109
- Microemulsions, 135–36
defined, 135
mixing of components, 135
- Millipede, 117
- Modulus of elasticity, 222
- Molecular beam epitaxy (MBE), 100
- Molecules
atomic bonding in, 26–35
four-atom, 31–32
many-atom, 34–35
six-atom, 32–34
three-atom, 31
two-atom, 29–30
- Multiwall CNTs (MWCNTs), 70, 173
- Nacre, 193–96
“brick-and-mortar” structure, 194
layer-by-layer (LbL) assembly, 195–96
- Nano biomaterials, 187–98
in biondiagnostics, 198–202
biomimetic nanotechnology, 188–98
in cell study, 202–6
introduction to, 187–88
- Nanocrystals, 44–47
- Nanocubes, 145
- Nanoelectronic materials, 153–84
CNT-based, 170–77
single-electron transistors (SETs), 153–70
spintronics, 178–84
- Nanografting, 117–19
benefits, 119
defined, 117
process schematic, 118
See also AFM lithography
- Nanograin-sized structural materials, 209–13
grain refinement, 210–13
thermal mechanical processing, 212–13
thermal method, 213
- Nanoimprint lithography (NIL), 110–14
acceptance, 110
air cushion press, 111–13
defined, 110
process, 110–11
schematic, 111
sequential embossing/imprinting lithography (SEIL), 113–14
3D patterning with, 111
- Nanoindentation, 115–17, 213–19
indentation hardness test, 213–14
load versus displacement curve, 215
principles of, 214
sample geometry, 215
in situ, 214–19
- Nanolithography, 107–26
AFM, 114–19
DPN, 119–20
introduction to, 107–10
NIL, 110–14
PPL, 119–22
templated self-assembly (TSA), 122–26
- Nanomaterials
bonding and band structures, 40–47
surface area, 69–70
surface crystallography, 55–57
surface energy, 57–61
surface reconfigurations, 61–68
surface science, 49–73
X-ray diffraction in, 78–80
- Nanomaterials characterization, 75–91
electron microscopy, 80–89
surface analysis methods, 89–91
X-ray diffraction, 75–80

- Nanoparticle-based detection methods, 200–201
- Nanoparticle-based SET, 167
- Nanoparticles
- CdS, 216–17
 - embedded, 169
 - equilibrium shapes, 59–61
 - inorganic, 139
 - isolated, 146
 - linear chains of, 146–47
 - mixing, 143
 - nonspherical, 142
 - nucleation, 70–71
 - of organic semiconductors, 138–40
 - self-assembly of, 140–47
 - surface layers and, 65–67
 - synthesis of, 129–40
 - volume, 70
 - Wulff constructions, 59–61
- Nanopatterning, 206
- Nanopen readers/writers, 118
- Nanoreactors, 138
- Nanoscale electronic mechanical system (NEMS), 217
- Nanostructural materials, 209–35
- mechanical instability, 219–35
 - nanograined-sized, 209–13
 - nanoindentation, 213–19
- Nanostructures
- in biondiagnostics, 198–202
 - in cell studies, 202–6
 - periodic array, buckling model, 229–35
 - schematic process flow, 199
- Nanotube-based detection methods, 201–2
- Nanowire-based detection methods, 201–2
- Negative voltage, 8
- 1D harmonic wave equation, 12–13
- 1D infinite potential well, 17–19
- quantization of wave function, 19
 - schematic illustration, 18
 - zero potential, 17
- 1D Schrödinger equations, 13–16
- defined, 14
 - rules, 14
 - time-independent, 15
- Orbitals
- defined, 22
 - highest occupied molecular (HOMOs), 32
 - linear combination of atomic, 26–35
 - lowest unoccupied molecular (LUMOs), 32
- Organic field effect transistor (OFET), 140
- Organic semiconductors, 138–40
- Overlap integral, 27
- Parallel replication, 108–10
- Pattern-dependent oxidation (PADOX), 166
- Patterned SET, 166–67
- Pauli exclusion principle, 23–24
- Peptide nucleic acid (PNA), 201
- Periodic array of nanostructures
- buckling model, 229–35
 - buckling process, 232
 - plane-strain problems, 232
 - potential energies, 232
 - rotation, 229
- Periodic table, 23–24
- Hund's rule, 24
 - Pauli exclusion principle, 23–24
- Phase separation, 168
- Phase transformation-based SET, 168–70
- Photoelectric effect, 4, 6–7
- Einstein explanation, 7–9
 - first observation, 6
 - illustrated, 6
 - rules, 6–7
 - work function from cathode metals from, 9
- Photolithography, 108
- Photons
- defined, 9
 - X-ray, 82
- Physical adsorption, 68, 69
- Physical vapor deposition (PVD), 100–102
- CNT fabrication, 173
 - sputtering, 101–2
 - summary, 105
 - thermal evaporation, 100–101
 - See also* Thin-film deposition
- Planck, Max, 4
- Planck's constant, 11
- Plane-strain problems, 232
- Plasma-enhanced CVD (PECVD), 104, 168
- Polymer pen lithography (PPL), 119–22
- defined, 119
 - force dependence, 121
 - maskless nature, 121
 - polymer pen array, 120

- process schematic, 121
- See also* Nanolithography
- Positive voltage, 8
- Potential well
 - electrons in, 17–21
 - 1D infinite, 17–19
 - 3D infinite, 19–21
- Pressure-temperature phase diagram, 137
- Quantum dots
 - defined, 43
 - energy bandgap, 43–44
 - ground state energy, 43
 - net charge in, 158
 - top-down method for, 41–44
- Quantum mechanics
 - beginning of development, 4
 - as driving force, 5
 - Heisenberg uncertainty principle, 5
 - history, 3–5
- Quantum mechanics effect, 163
- Quantum tunneling, 87
- Quantum wells
 - creating, 41
 - defined, 41
 - energy bandgap, 41
 - top-down method for, 41–44
- Reactive ion etching (RIE), 110
- Reprecipitation method, 139
- Reverse micelle model, 136
- Scanning electron microscope (SEM),
 - 86–87, 112
- Scanning probe lithography (SPL), 107, 114
- Scanning probe microscope (SPM), 87–89
 - constant current image, 88
 - defined, 87
 - high-resolution topographs, 89
 - illustrated, 88
 - operating mechanisms, 87–88
 - quantum tunneling, 87
- Scanning tunneling microscope (STM), 87
- Schrödinger equations, 12–16
 - free electrons and, 16
 - 1D, 13–16
 - time independent, 12–16
- Secondary electrons, 82
- Secondary ion mass spectroscopy (SIMS), 91
- Self-assembled monolayer (SAM), 68, 117, 144
- Self-assembly-based SET, 167
- Self-assembly of nanoparticles, 140–47
 - collective properties, 146–47
 - electrostatic, 141–42
 - hydrogen bonding-based, 140–41
 - hydrophobic, 144
 - shape-selective, 142–44
 - template-assisted, 144–46
- Semiconductors
 - energy band for, 38–40
 - “gapless,” 171
 - organic, nanoparticles of, 138–40
- Serial writing, 110
- Shape-selective assembly, 142–44
- Shells, 193, 194
- Single-electron capacitors, 154–58
- Single-electron transistors (SETs), 153–70
 - advantages of, 168
 - defined, 158
 - drain, 158, 160
 - equivalent circuit, 160
 - fabrication of, 165–70
 - gate electrode, 158
 - illustrated, 159
 - nanoparticle-based, 167
 - “off” state, 162
 - operating conditions, 162
 - operating principles, 158–62
 - PADOX, 166
 - patterned, 166–67
 - phase transformation-based, 168–70
 - quantum effect on, 162–65
 - self-assembly-based, 167
 - single-electron capacitor, 154–58
 - source, 158, 160
- Single wall CNTs (SWCNTs), 70
 - diffraction pattern, 85
 - grown by CVD method, 174
 - high yields, 173
 - VLS mechanism for formation of, 174
 - See also* Carbon nanotubes (CNTs)
- Six-atom molecule (benzene ring), 32–34
- Sol-gel process, 132–35
 - defined, 132
 - elimination types, 134
 - four-coordinated tetraethoxy, 132
 - typical steps, 133
- Solid parallel-plate press (SPP), 113
- Solvothermal process, 137

- Spin field emission transistor (SFET),
 - 181–82
 - challenges, 182
 - Datta-Das, 181
 - types of, 182
- Spin magnetic momentum, 178
- Spin torque transfer (STT), 183
- Spintronics, 178–84
 - defined, 178
 - devices, 180–84
 - giant magnetoresistance effect, 178–80
 - GMR storage device, 180–81
 - magnetoresistive random access memory (MRAM), 182–84
 - operating principles, 178–80
 - spin field emission transistor (SFET), 181–82
 - spin magnetic momentum, 178
- Sponge spicules, 193
- Sputtering PVD process, 101–2
- Step propagation, 96–97
- Stepwise contraction and adsorption
 - nanolithography (SCAN), 203–6
 - concept illustration, 204
 - defined, 203
 - dimensions, 206
 - minimization principle, 203
 - 1D process, 206
- Stranski-Krastanov model, 98–99
- Structural biomimicry, 191–98
 - compound eyes, 196–98
 - gecko foot, 191–93
 - nacre, 193–96
- Supercritical fluids (SCFs), 137
- Supercritical process, 137
- Surface analysis methods, 89–91
 - Auger electron spectroscopy (AES), 90
 - secondary ion mass spectroscopy (SIMS), 91
 - X-ray photoelectron spectroscopy (XPS), 90–91
- Surface area, 69–70
- Surface atoms, percentage of, 66
- Surface crystallography, 55–57
- Surface energy, 57–61
 - crystallographically preferred surface, 59
 - defined, 58
 - estimating with Wulff constructions, 59–60
- Surface reconstructions, 63–65
 - relaxation, 62–63
 - surface layers and nanoparticles, 65–67
- Surface reconstruction, 63–65
 - defined, 62
 - illustrated, 64
 - Wood's notation, 64–65
- Surface relaxation
 - defined, 62
 - lateral, 63
 - vertical, 62–63
- Surface symmetry, 57
- Synthesis of nanoparticles, 129–40
 - coprecipitation, 130–32
 - hydrothermal/solvothermal methods, 136–37
 - metal-oxide, 134
 - microemulsions, 135–36
 - organic semiconductors, 138–40
 - sol-gel process, 132–35
 - templated synthesis, 138
- Template-assisted assembly, 144–46
- Templated self-assembly (TSA), 122–26
 - of block copolymers, 122–26
 - in defect elimination, 123
 - defined, 122
 - spherical colloid assembly, 124
 - system illustration, 123
- Templated synthesis, 138
- Thermal evaporation, 100–101
- Thermal mechanical processing, 212–13
- Thermodynamic equilibrium state, 70
- Thin-film deposition, 95–105
 - CVD, 102–5
 - heterogeneous film growth mechanisms, 97–99
 - homogeneous film growth mechanisms, 95–97
 - mechanisms, 95–99
 - methods, 99–105
 - PVD, 100–102
- Thin films
 - buckling, 220
 - wrinkling, 219–23
- 3D infinite potential well, 19–21
 - ground state of energy, 20
 - Schrödinger equation, 19
 - second-level state, 20–21

- total energy of electron, 20
- zero potential, 19
- Three-atom molecule, 31
- Time independent Schrödinger equation, 12–16
- Timoshenko's theory of elastic stability, 222
- TiN/MgO thin films, 217–19
- Tissue engineering, 205, 206
- Tooth, 193, 194
- Top-down method
 - defined, 40
 - nanolithography, 107–26
 - for quantum wells and dots, 41–44
 - thin-film deposition, 95–105
 - See also* Bottom-up method
- Transmission electron aberration-corrected microscope (TEAM), 84
- Transmission electron microscope (TEM), 83–86
 - high-resolution (HRTEM), 84, 134, 169
 - nanindentation tests with, 215
 - schematic illustration, 83
 - in situ, 174
 - uses, 84–86
 - See also* Electron microscopy
- Transmitted electrons, 82
- Tunneling
 - illustrated, 163
 - quantum, 87
 - as quantum mechanical effect, 163
- Two-atom molecule, 29–30
- Valance band (VB), 36–37, 40
- Vertical surface relaxation, 62–63
- Volmer-Weber model, 97–98
- Wetting, 71–73
 - angle, 72
 - complete, 72–73
 - defined, 71
 - dewetting and, 73
 - partial, 72–73
 - promoting, 72
- Wood's notation, 64–65
- Wrinkling of thin films, 219–23
 - geometry, 223
 - illustrated, 221
 - patterns, 220
 - phenomenon, 222
 - problem solving, 222–23
 - top surface, 219–20
- Wulff constructions, 59–61
 - defined, 60
 - illustrated, 61
 - surface energy estimation with, 59–60
- X-ray diffraction, 75–80
 - Bragg's law and, 77–78
 - Laue method and, 76–77
 - in nanomaterials, 78–80
 - patterns, 79
- X-ray photoelectron spectroscopy (XPS), 90–91
- X-rays, 82
- X-ray scattering experiment, 10
- Young's equation, 72
- Zeeman effect, 178